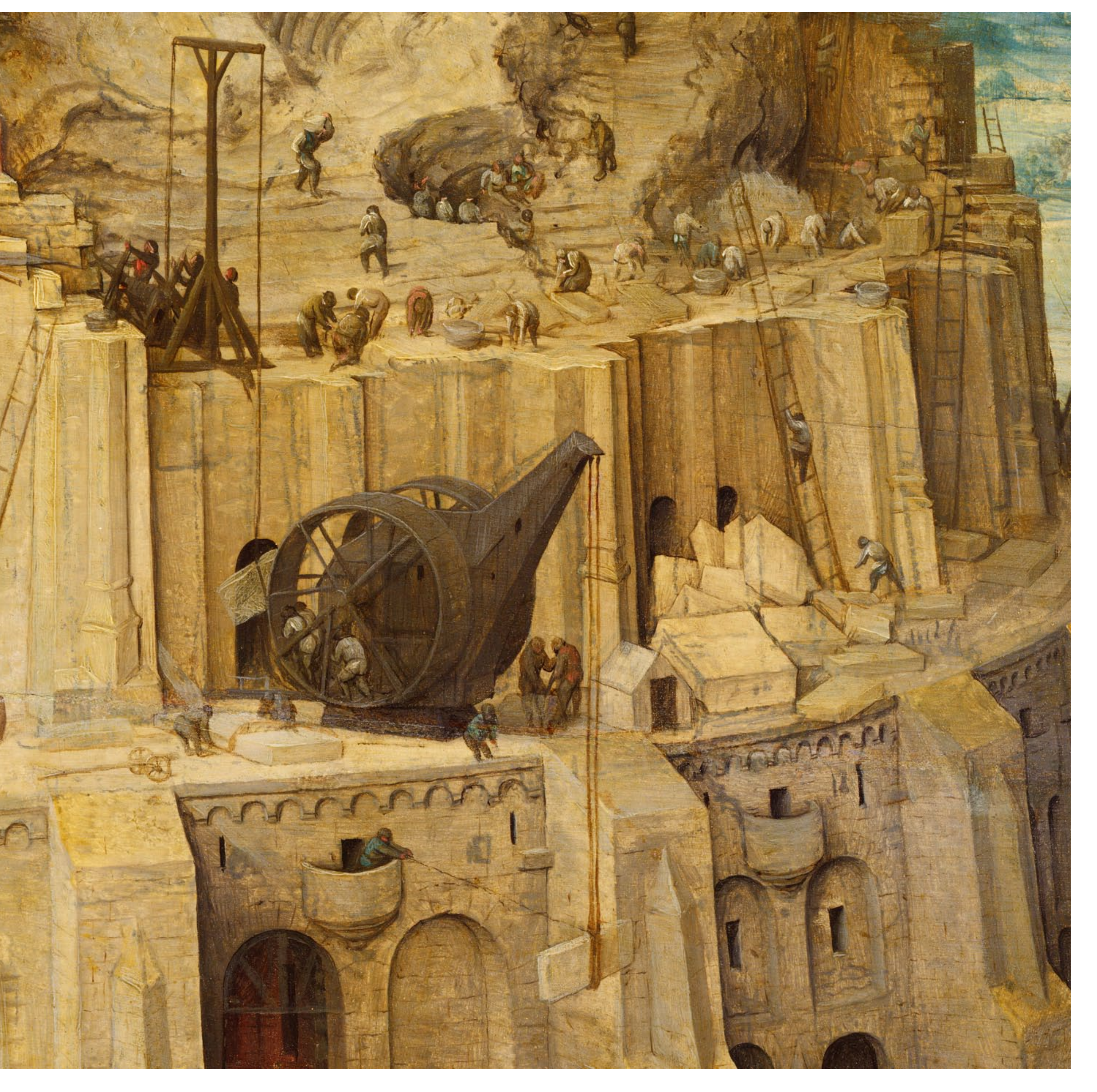
INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**



TECNOLOGIA DE BRUEGEL

Mayo 1978

170 Ptas.

Número 20
Mayo 1978



EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-29
ESPAÑA

Colaboradores de este número:

Asesoramiento y traducción:

Joan Albaigés: *La producción mundial de petróleo;* Pedro Pascual: *Leptones pesados;* Manuel Puigcerver: *La superficie de Marte;* Joandomènec Ros: *El flujo de energía en un ecosistema de bosque;* Carles Oppenheimer: *Cómo fabrican ATP las células;* Luis Bou: *Matemática combinatoria de la planificación;* Julio Samsó: *Bruegel el Viejo, guía para el estudio de la ciencia del s. XVI;* Luis Bou: *Juegos matemáticos;* José E. Myro y Alicia Menéndez: *Taller y laboratorio;* Vicente Maraña (C. Invest. CTNE): *Ciencia y sociedad.*

Ciencia y sociedad:
Peter P. Luff

Libros:
Ramón Margalef y Ramón Bayés



LA PORTADA

La fotografía de la portada muestra un detalle del cuadro *La Torre de Babel,* del artista flamenco del siglo XVI, Pieter Bruegel (véase "Bruegel el Viejo, guía para el estudio de la ciencia del s. XVI" por H. Arthur Klein, en este número). Se trata de una grúa capaz de levantar objetos bastante pesados, cuyo mecanismo elevador está accionado por hombres instalados en el interior de un molino. Hay pruebas de que estas grúas se empleaban en el siglo XVI en la construcción y en la carga y descarga de barcos. Como puede verse en una reproducción más amplia que aparece en el artículo, esta grúa se halla en una rampa y se emplea para levantar materiales. Las pinturas de Bruegel se conservan hoy en el Kunsthistorisches Museum de Viena.

Suscripciones:
Prensa Científica, S.A.
Departamento de suscripciones
Calabria 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Distribución para España:
Distribuciones de Enlace, S.A.
Ausias March, 49
Barcelona-10 (España)

Distribución para los restantes países:
Editorial Labor, S.A.
Calabria, 235-239
Barcelona-29 (España)

Precio de venta de este número:
España: 170 pesetas
Extranjero: 2 U.S. $

Condiciones de suscripción:
España:
Un año (12 números): 1870 pesetas
Extranjero:
Un año (12 números): 30 U.S. $

Ejemplar atrasado ordinario:
190 pesetas

Ejemplar atrasado extraordinario:
270 pesetas.

Difusión controlada por **OJD**

PROCEDENCIA DE LAS ILUSTRACIONES

Grabado de la portada por cortesía del Kunsthistorisches Museum de Viena

Página	Fuente
7-3	Alan D. Iselin
17	Stanford Linear Accelerator Center
18-24	Gabor Kiss
27-36	National Aeronautics an Space Administration
36	*Arachnida.* © 1977, Academic Press
37	Ilil Arbel
38	National Aeronautics an Space Administration
41-42	Adolph E. Brotman
46	Robert L. Talerico, U.S. Forest Service
47-53	Tom Prentiss
54	Robert L. Talerico, U.S. Forest Service
55-56	Tom Prentiss
59	John N. Telford, Cornell University
62-74	Bunji Tagawa
77-84	Ilil Arbel
87	Kunsthistorisches Museum, Viena
88	Museo de Capodimonte, Nápoles
89	Groeningemuseum, Brujas *(arriba);* Devonshire Collection, Chatsworth, reproducido con permiso de Trustees of the Chatsworth Settlement *(abajo)*
90-92	Mr. y Mrs. Jacob I. Zeitlin
95-96	Antonio Fontdevila
97	Anthony C. Allison
98	Antonio Fontdevila
99	*Scientific American*
100-102	Antonio Fontdevila
105-109	Ilil Arbel
110-116	Michael Goodman

Dep. legal: B. 38.999-76
Fotocomposición: Fort, S.A.
Rosellón, 33 - Barcelona-29
Cayfosa. Santa Perpetua de Moguda. Barcelona.

Printed in Spain - Impreso en España

Cartas

Señor Director:

Tal como George M. Woodwell observa ("La cuestión del dióxido de carbono", INVESTIGACION Y CIENCIA, marzo, 1978) ninguna forma de control puede permanecer insensible al "efecto invernadero". Utilizando como guía nuestro limitado conocimiento actual, el dióxido de carbono es un problema de primer orden, con posibles consecuencias geopolíticas en un plazo de 25 años. Quizá sean de utilidad algunos comentarios adicionales sobre el tema.

Durante los últimos 117 años, la producción de dióxido de carbono ha aumentado exponencialmente a un ritmo del 4,25 por ciento anual; aumento que se interrumpió tan sólo por las dos guerras mundiales y por la Gran Depresión, que disminuyeron nuestra tasa de producción de dióxido de carbono durante 23 años (quién sabe si concediéndonos tiempo para reconsiderar nuestra política energética). Yo remarcaría que ninguna discontinuidad de menor escala en los modelos sociales ha tenido efecto alguno sobre la tasa de crecimiento: el advenimiento de la energía nuclear, el embargo del petróleo y otros cambios voluntarios que puedan haber ocurrido no han tenido trascendencia en los datos.

La tendencia atmosférica puede ser representada con precisión por una ecuación en la que los parámetros de crecimiento corresponden a los datos de combustibles fósiles y en la que se considera que el 39 por ciento de la producción total de dióxido de carbono va a parar al "océano". Dentro de la precisión de estos datos, la contribución de las plantas terrestres es nula o proporcional al combustible fósil (con un porcentaje diferente de dióxido de carbono que pasa al océano).

No existe ningún problema en disponer del océano para el dióxido de carbono "extraviado", si el modelo no se limita a la mezcla por difusión en remolino. El transporte biológico por las diurnas migraciones verticales de los organismos de la "capa profunda de dispersión" podría transportar fácilmente todo el dióxido de carbono antropogénico a través de la termoclina. Un modelo tal recalca la unión del dióxido de carbono con otros nutrientes, especialmente con el fósforo. Sin embargo, si la unión es fuerte, la cantidad del dióxido de carbono separado de la atmósfera de esta forma tiene un límite superior constante con el tiempo y, por tanto, dicha cantidad no aumentaría en función del incremento del dióxido de carbono atmosférico. En este caso, avanzamos hacia el problema más rápidamente de lo que cualquier modelo predice (ya que éstos consideran que la cantidad separada es proporcional a la presente).

La única posibilidad que, en mi opinión, puede ofrecernos un mayor respiro es la de que la unión sea débil, con lo que el carbono (metabolizado para la obtención de energía) y el fósforo (metabolizado para el crecimiento) son liberados en diversos momentos del curso de un organismo y, por tanto, son liberados a distintas profundidades marinas.

El desacuerdo potencial en la capacidad de almacenamiento y el hecho de que el efecto invernadero sea obra exclusiva del hombre (pudiendo evitarse si así lo decidiera) se combinan para sugerir que ha llegado la hora de acometer, interdisciplinariamente, el estudio de la cuestión del dióxido de carbono.

FERREN MACINTYRE

Profesor visitante de Oceanografía
Universidad de Rhode Island
Kingston

Señor Director:

En el artículo "La cuestión del dióxido de carbono" del número de Investigación y Ciencia correspondiente a marzo de 1978, el profesor G.M. Woodwell, al especular sobre los peligros potenciales que puede originar un aumento de la concentración del CO_2 en la atmósfera, parece no tener en cuenta que el complejo proceso fotosintético se puede descomponer en dos fases fundamentales.

Una, en la que el H_2O absorbido por las plantas es descompuesto por los cuantos de luz de la energía luminosa en oxígeno molecular e hidrógeno atómico y, otra, en la que el hidrógeno, que se ha combinado con un aceptor apropiado y ha quedado almacenado en la planta, se emplea en la transformación del CO_2 en azúcares.

Pues bien, tres son los factores externos de primordial importancia en la regulación de la actividad fotosintética: la intensidad de la luz, la concentración de CO_2 en el aire y la temperatura, y cualquiera de los tres puede actuar como factor limitante. Así, la fotosíntesis resulta limitada por la cantidad de luz aprovechable cuando la intensidad luminosa es débil, no pudiendo, en estas condiciones, incrementarse el fenómeno aumentando la concentración de CO_2 o la temperatura. En cambio, para una intensidad luminosa alta la actividad fotosintética está limitada por la cantidad de CO_2 utilizable y, por tanto, puede aumentarse elevando la concentración por encima de los valores normales del 0,03 por ciento.

Por otro lado, el valor de la fotosíntesis se intensifica apreciablemente elevando la temperatura cuando el factor limitante es el CO_2 y no la luz.

Por tanto, si la desforestación produce un incremento positivo en la concentración de CO_2 en el aire, y el consiguiente aumento de absorción calorífica por la atmósfera terrestre, habrá de calcularse si la actividad fotosintética, al verse favorecida por el aumento de CO_2 y la elevación de la temperatura, será capaz de mantener en equilibrio las concentraciones de los componentes atmósfericos. Es decir, ¿será capaz una masa forestal menor a la actual de realizar una actividad fotosintética igual a la que se da ahora, apoyándose en una mayor concentración de CO_2 y en una temperatura más elevada?

Dr. ALFRED A. ALLENDE YOHN

Colaborador Científico del C.S.I.C.
C.E.B.A.S.
Murcia

ERRATA

Número 17, febrero de 1978

Pág.	*columna*	*línea*	*dice*	*debe decir*
11	3	29-30	anticuerpo y antitubulina-tubulina	anticuerpo antitubulina-tubulina
11	3	33-34	fluorescente fluoresceína	floruro fluorescente fluoresceína
12	2	20	involucionando	despegándose

Los autores

ANDREW R. FLOWER ("La producción mundial de petróleo") es analista *(policy analyst)* de la British Petroleum Company, especializado en suministro y demanda energética a largo plazo. Natural de Yorkshire, terminó su licenciatura en matemáticas en la Universidad de Oxford en 1968 y recibió el grado de master en estadística en la Universidad de Birmingham al año siguiente. Entró luego en la British Petroleum como experto en estadística, dedicándose durante tres años al estudio de aplicaciones de los métodos prospectivos para la planificación de procesos de la compañía. En 1972 pasó al Departamento de Planificación Corporativa como economista responsable de la prospección de la demanda de energía en la Comunidad Económica Europea; en 1975 se le asignó su actual empleo en la Policy Review Unit.

MARTIN L. PERL y WILLIAM T. KIRK ("Leptones pesados") son ambos miembros del Stanford Linear Accelerator Center (SLAC). Perl, profesor de física en la Universidad de Stanford, se graduó en ingeniería química en el Polytechnic Institute de Brooklyn en 1948. Sus intereses cambiaron hacia la física atómica, y en este campo se doctoró por la Columbia University en 1955. Al entrar a formar parte de la facultad de la Universidad de Michigan empezó a trabajar en física de partículas elementales y, desde entonces, ha permanecido en este campo. En el SLAC el trabajo de Perl se centra en la física de los anillos de colisión electrón-positrón con la meta de lograr entender mejor los leptones. Kirk, escritor de artículos técnicos, es director asistente de la división de investigación del SLAC y editor de la publicación mensual del laboratorio, *SLAC Beam Line.* Se graduó en historia en la Cornell University en 1952 y, después de cuatro años en la industria, entró a formar parte de los W. W. Hansen Laboratories of Physics de Stanford.

RAYMOND E. ARVIDSON, ALAN B. BINDER y KENNETH L. JONES ("La superficie de Marte") son geólogos con un interés especial en los planetas. Arvidson, que es profesor adjunto de ciencias terrestres y planetarias en la Universidad de Washington y director del Grupo de Imágenes de los Módulos de Descenso Viking, recibió su doctorado en Ciencias de la Tierra en la Universidad Brown en 1974. Su trabajo sobre la superficie de Marte, escribe, "ha despertado interés sobre las regiones desérticas de la Tierra y la manera como se reducen y extienden, particularmente en regiones donde el crecimiento de los desiertos es económicamente devastador". Binder, científico investigador en el Instituto de Geofísica de la Universidad de Kiel en Alemania Occidental, obtuvo su doctorado en la Universidad de Arizona. En la actualidad se ocupa en el análisis de los datos lunares y del Viking y en "defender la teoría de la formación de la Luna por fisión de la Tierra". Jones, investigador asociado en Brown y miembro del grupo Viking, obtuvo su doctorado en Brown en 1974.

JAMES R. GOSZ, RICHARD T. HOLMES, GENE E. LIKENS y F. HERBERT BORMANN ("El flujo de energía en un ecosistema de bosque") son biólogos especializados en la ecología y la gestión de bosques. Gosz, profesor asociado de biología en la Universidad de Nuevo México, obtuvo su doctorado en ciencia forestal por la Universidad de Idaho en 1968. En su palmarés investigador hay numerosos estudios sobre el impacto de las actividades humanas en los sistemas naturales. Holmes, profesor de biología en el Dartmouth College, obtuvo su doctorado en zoología por la Universidad de California en Berkeley en 1964. Su área de interés es la estructura y función de las comunidades animales. Likens, profesor de ecología y sistemática en la Universidad de Cornell, obtuvo su doctorado en zoología por la Universidad de Wisconsin en 1962. Su investigación se centra en el ciclo de nutrientes y minerales en ecosistemas lacustres y forestales. Bormann, profesor de ecología forestal en el School of Forestry and Environmental Studies de la Universidad de Yale, obtuvo su doctorado en ecología vegetal por la Universidad de Duke en 1952.

PETER C. HINKLE y RICHARD E. McCARTY ("Cómo fabrican ATP las células") son profesores de la Universidad de Cornell. Hinkle, que enseña biología molecular y celular, se graduó en el Harvard College. Tras doctorarse por la Universidad de Nueva York pasó a investigar la transformación biológica de la energía con Peter Mitchell en los Glynn Researh Laboratories de Inglaterra. Su trabajo actual se enfoca hacia cómo diversas substancias, particularmente el azúcar glucosa, son transportadas a través de las membranas celulares. McCarty, que enseña también biología molecular y celular, estudió en Cornell, obteniendo su doctorado por la Universidad de Johns Hopkins en 1964.

RONALD L. GRAHAM ("Matemática combinatoria de la planificación") es director del departamento de matemática discreta de los Laboratorios Bell. Recibió su título de bachiller en física por la Universidad de Alaska y su grado de doctor por la Universidad de California en Berkeley, en 1962. Su trabajo actual implica teoría combinatoria, teoría de números, algoritmos y la aplicación de estas teorías a los problemas del mundo real.

H. ARTHUR KLEIN ("Pieter Bruegel el Viejo como guía para el estudio de la ciencia del s. XVI") reside en Malibu, California. Ha escrito, colaborado o dirigido más de veinte libros, relacionados fundamentalmente con la ciencia o con las Bellas Artes. El más reciente es una obra de 740 páginas titulada *The World of Measurements* (Simon and Schuster, 1974).

ANTONIO FONTDEVILA ("El mantenimiento de la variabilidad genética de las poblaciones") es especialista en genética evolutiva de *Drosophila.* Después de doctorarse en 1971 con el Profesor Antonio Prevosti en la Universidad de Barcelona, trabajó durante dos años en la genética de los polimorfismos enzimáticos con los profesores Theodosius Dobzhansky y Francisco J. Ayala en la Universidad de California. En 1973 obtuvo la cátedra de Genética que ocupa actualmente en la Universidad de Santiago de Compostela. Desde entonces ha realizado diversos trabajos de genética y biología de poblaciones en dicha Universidad y también en las Universidades de Arizona y Hawaii, con los profesores William B. Heed y Hampton L. Carson, respectivamente. Durante sus estancias en los Estados Unidos ha pronunciado conferencias en distintas universidades norteamericanas. En 1975 fue elegido por el British Council para visitar universidades británicas, donde pronunció también una serie de conferencias. Es autor de diversos trabajos de investigación publicados en revistas nacionales y extranjeras.

La producción mundial de petróleo

Las existencias de petróleo en el subsuelo son limitadas, al igual que su tasa de recuperación. Antes de alcanzar el año 2000, el menguante suministro de crudos no podrá satisfacer su creciente demanda

Andrew R. Flower

El suministro de crudos de petróleo será incapaz de hacer frente a su creciente demanda antes del año 2000. Puesto que la producción del petróleo tiende inevitablemente a estabilizarse, para después empezar a descender, la demanda energética, que continuará creciendo a pesar de los fuertes intentos por controlarla, deberá satisfacerse con otros combustibles. Los países importadores de crudos tienen sólo cinco años o quizá veinte, como mucho, para pasar de depender del petróleo a confiar en otros combustibles fósiles, en la energía nuclear y, eventualmente, en fuentes renovables de energía. Puesto que para el desarrollo de nuevas fuentes energéticas se requieren grandes inversiones y tiempo, los esfuerzos deben iniciarse inmediatamente.

Estas son, entre otras, las conclusiones más importantes de un estudio internacional de dos años de duración, el Seminario sobre Estrategias Energéticas Alternativas. Alrededor de 70 personas del mundo de los negocios y de la industria, del gobierno y de las universidades, pertenecientes a 15 importantes países no comunistas e importadores de petróleo, se reunieron bajo la dirección de Carroll L. Wilson, del Instituto de Tecnología de Massachusetts, para estudiar el abastecimiento y la demanda mundiales de energía, detectar posibles problemas en la satisfacción de las necesidades energéticas y considerar las estrategias que deberían acometerse en la resolución de dichos problemas. (El ámbito geográfico considerado fue el que se encuentra fuera del área comunista. La URSS y China son importantes productores y consumidores de energía, pero las estadísticas sobre sus reservas, producción, consumo y comercio, en el mejor de los casos, son imprecisas y su comercio con los países no comunistas no parece importante.) Los participantes en el Seminario procedían de Estados Unidos, Canadá, México, Venezuela, Gran Bretaña, Francia, Alemania Federal, Holanda, Dinamarca, Suecia, Noruega, Finlandia, Italia, Irán y Japón.

El Seminario desarrolló proyectos de suministro y demanda de energía bajo diversos supuestos, estrechamente relacionados con el crecimiento económico y el coste de la energía. Se formularon dos modelos: el primero implicaba una elevada tasa de crecimiento económico y elevados precios energéticos; el segundo, una tasa baja de crecimiento y precios constantes. El modelo de alto crecimiento y precios elevados supuso que el crecimiento económico en el mundo no comunista sería, por término medio, del 5,2 por ciento anual entre 1976 y 1985, y del 4 por ciento entre 1985 y el año 2000, y que los precios energéticos permanecerían constantes (en términos reales) hasta 1985, con un incremento del 50 por ciento hacia el año 2000. El modelo de bajo crecimiento y de precios constantes supuso que el crecimiento económico medio sería del 3,4 por ciento anual hasta 1985 y del 2,8 por ciento hasta el 2000, y que los precios reales energéticos permanecerían constantes durante la totalidad del período.

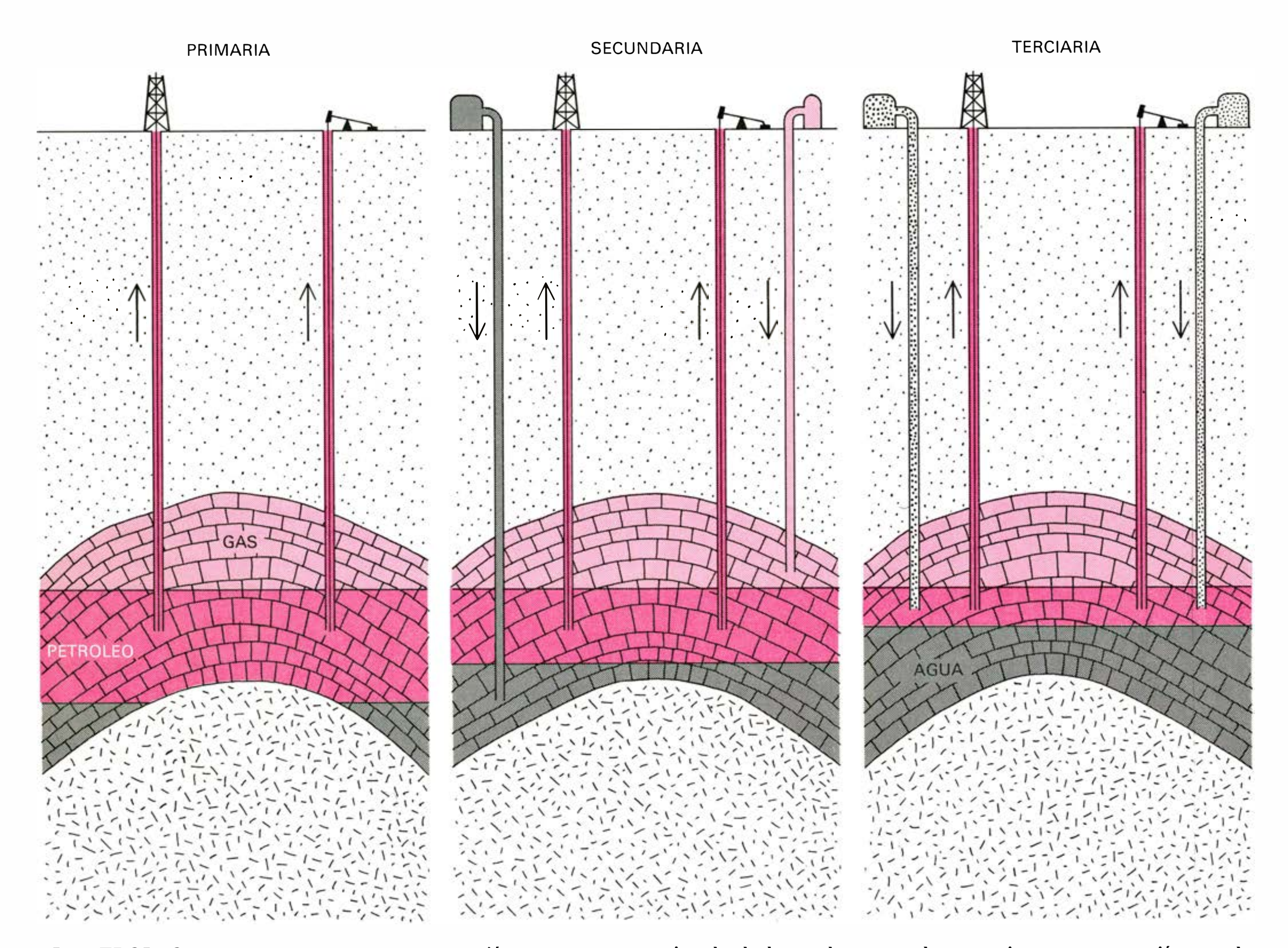
Dentro de este esquema se prepararon las previsiones regionales de la demanda energética para cada combustible y se sumaron al objeto de planear la demanda energética total en 1985 y en el año 2000. En el caso del petróleo, que hoy satisface más de la mitad de las necesidades energéticas de las sociedades industriales, los planes indican que los refuerzos de conservación de energía tendrán algún efecto, de forma que el crecimiento de la demanda será menor que el 6 por ciento medio anual entre 1960 y 1972. Supuesto un elevado crecimiento económico y unos precios del petróleo altos, se postuló un incremento de la demanda de crudos del 3,6 por ciento anual hasta 1985 y del 2,6 por ciento anual entre 1985 y el 2000. Con un ritmo de crecimiento bajo y unos precios constantes, los supuestos incrementos de la demanda del petróleo fueron del 2,5 por ciento y del 1,8 por ciento, respectivamente. Estos índices determinaron dos curvas para la demanda de petróleo. El paso siguiente consistió en considerar el tiempo durante el cual el suministro de crudos satisfaría aquellas demandas. (Aunque el Seminario trataba principalmente de las previsiones energéticas hasta el año 2000, la importancia central del petróleo nos llevó a considerar las perspectivas de las reservas y la producción petrolífera hasta el 2025; estas últimas son, sin embargo, altamente especulativas.) En este artículo trataré de cómo el Seminario proyectó las estimaciones del futuro suministro de crudos y alcanzó sus conclusiones en cuanto a las relaciones entre suministro y demanda.

A menudo se cree, de forma errónea, que el petróleo se halla en extensas cavidades subterráneas, que sólo necesitan ser bombeadas hasta la sequedad para recuperarlo totalmente. En realidad, el petróleo se encuentra en los pequeños huecos de rocas porosas, algo así como el agua en una esponja. A través del material poroso migra lentamente, hasta que es atrapado por una roca impermeable, formando generalmente una capa entre agua y gas. El petróleo se encuentra normalmente sometido a presión,

y la perforación de un pozo descarga dicha presión; el petróleo fluye hacia el pozo por el que asciende hasta la superficie, espontáneamente (si la presión en el campo es suficiente) o por bombeo. La recuperación del petróleo a presión natural se conoce como la "producción primaria", que hoy día en los Estados Unidos significa por término medio la recuperación del 25 por ciento del petróleo existente en un yacimiento. La "recuperación secundaria" se consigue, en los yacimientos en que es posible, bombeando agua o gas hasta el interior de los mismos, para incrementar la presión del petróleo. La "recuperación terciaria", todavía no aplicada ampliamente, puede lograrse, a veces, disminuyendo la viscosidad del petróleo de forma que fluya más fácilmente, bien por calentamiento (inyectando vapor, por ejemplo) o inyectando productos químicos adecuados. En los Estados Unidos, los métodos de extracción secundaria y terciaria han incrementado la recuperación del petróleo contenido en los yacimientos desde un 25 por ciento en los años 40 hasta 32 por ciento en la actualidad, aproximadamente. La mayor parte del petróleo se produce en el mundo en campos convencionales por estos métodos y por la recuperación de los gases licuados, como productos secundarios del gas natural. El Seminario sólo consideró estas fuentes, excluyendo el petróleo de los esquistos, pizarras bituminosas y licuefacción del carbón.

Los principales factores que determinan el posible suministro de crudos son la cantidad de reservas comprobadas (petróleo recuperable de las reservas conocidas, a precios normales y con tecnología corriente) y el ritmo al cual aquellas reservas pueden ser explotadas. Cada campo tiene un ritmo potencial de producción que depende de su extensión, geología e instalaciones disponibles. En general, es imposible obtener una producción anual de más del 10 por ciento de las reservas recuperables en un campo sin reducir la cantidad de petróleo que puede ser finalmente recuperada. En algunos campos, el ritmo puede ser alto y en otros bajo; con todo, una relación de reservas-producción *(R/P)* de 10 a 1 es probablemente la mínima posible para las reservas mundiales de petróleo. Aplicar globalmente esta relación de 10:1 implicaría, sin embargo, que todos los campos de petróleo conocidos podrán producir, a la vez, al máximo ritmo. La realidad es que siempre habrá algunos campos en desarrollo y no serán, por tanto, productivos, a pesar de contribuir al total de las reservas comprobadas. Una estimación más razonable del ritmo máximo de producción de las reservas existentes puede venir dado por una relación *R/P* de 15 a 1.

Lo que se necesita para proyectar



EL PETROLEO se encuentra, generalmente, a presión en rocas porosas y en estructuras geológicas. La perforación de un pozo hasta la formación que contiene el petróleo descarga la presión y, en la producción primaria, el crudo fluye hasta la superficie, espontáneamente *(a)* o por bombeo *(b)*. Cuando la presión natural es demasiado baja para impulsarlo hasta el pozo, puede conseguirse una recuperación secundaria bombeando agua *(c)* o gas *(d)* dentro del campo a fin de incrementar la presión. En la recuperación terciaria, todavía no aplicada ampliamente, se disminuye la viscosidad del petróleo inyectando vapor, que lo calienta *(e)*, o bien inyectando algún producto químico adecuado *(f)*.

una curva base de suministro de petróleo es, por consiguiente, una gráfica anual de las reservas existentes comprobadas. Esta se obtiene añadiendo, a las reservas comprobadas de un año determinado, las nuevas y restando la cantidad de petróleo que ha sido extraído, es decir, la producción real acumulada. La máxima producción potencial de tales reservas se determina, entonces, dividiéndolas por 15. En el momento en que la demanda de crudos es menor que la producción potencial, la producción real iguala por definición a la demanda; cuando la demanda excede la producción potencial aquélla está necesariamente limitada a ésta, que se convierte entonces en la producción real. La cuestión importante que se plantea es: ¿en qué punto una relación *R/P* de 15 a 1 sitúa la producción potencial por debajo de la demanda proyectada? Para determinar este dato (habiendo estimado la demanda) necesitamos conocer las reservas comprobadas en un año base y el índice probable anual de incremento bruto de las reservas.

Las estimaciones de reservas recuperables comprobadas son inciertas, y varían de un año para otro. El análisis del Seminario se basó en los datos publicados a finales de 1975 por el *Oil and Gas Journal*, recopilados a partir de los cálculos efectuados por los gobiernos y las compañías petrolíferas. Las reservas existentes comprobadas se calcularon entonces en 555.000 millones de barriles, de los cuales un 80 por ciento en los países miembros de la Organización de Países Exportadores de Petróleo (OPEP), la mayoría en el Oriente Medio. Por otra parte, la producción se ha desarrollado principalmente hasta la fecha en Norteamérica. En total, alrededor de 846.000 millones de barriles se habían descubierto en el mundo no comunista hasta 1975, de los cuales se había consumido aproximadamente una tercera parte.

Una guía del progreso de los descubrimientos en el futuro lo constituye el ritmo al cual aumentaron las reservas en el pasado. Hay dos formas de evaluarlo. Se puede simplemente comparar los datos publicados cada año o se pueden pasar las modificaciones (generalmente incrementos) de las estimaciones del petróleo existente en cada campo al año en el cual tuvo lugar el descubrimiento. Por ejemplo, un campo descubierto en 1960, al que inicialmente podrían habérsele atribuido unas reservas recuperables de mil millones de barriles, a los cinco años, con la información obtenida durante el desarrollo del mismo o por las mejoras en las técnicas de recuperación, pueden evaluarse de nuevo las reservas, cifrándolas en 1200 millones de barriles. Las estimaciones anuales atribuirían entonces estos doscientos millones de más a 1965 y en el segundo caso, en cambio, se computarían con los descubrimientos del año 1960.

El método de año por año pone de relieve los cambios corrientes en el índice de aumento de las reservas, por las nuevas evaluaciones de campos ya conocidos, así como por la incidencia de los descubrimientos. En la ilustración superior de la página siguiente se muestra cómo la media quinquenal (que suaviza las fluctuaciones anuales) se eleva desde alrededor de 22.000 millones de barriles por año, entre 1950 y 1965, hasta más de 50.000 millones, entre 1965 y 1970, cuando fue reconocida la magnitud de las reservas existentes en los campos de Oriente Medio, descubiertos antes de 1950. Desde 1970, los incrementos anuales de las reservas han sido sólo de alrededor de 25.000 millones. Por otra parte, el método de asignar las nuevas evaluaciones de los campos al año en que fueron descubiertos pone de relieve que la mayor parte del incremento de finales de los años 60 fue debida a los campos descubiertos antes de 1950 *(véase la ilustración inferior de la página siguiente).* Dicho método nivela el índice entre 1950 y 1970 a una media de 18.000 millones de barriles por año, con un descenso hasta 15.000 millones a partir de 1970.

Al efectuar estimaciones para el futuro es necesario considerar dos componentes diferentes en el aumento bruto de las reservas: primero, las reservas correspondientes exclusivamente a los nuevos descubrimientos y, segundo, los incrementos de las reservas de los campos descubiertos antes de 1975, tanto por una nueva evaluación de las mismas como por la mejora de las técnicas de recuperación del petróleo.

La tasa de crecimiento en el pasado de las nuevas reservas, propiamente dichas, hubiese sido más baja sin el descubrimiento de las reservas masivas de Oriente Medio, donde han sido halladas en una región de tan sólo 1300 por 800 kilómetros, aproximadamente, alrededor del 60 por ciento de todas las reservas conocidas del mundo no comunista. ¿Podría encontrarse de nuevo una región tan rica en petróleo? No parece muy probable. Muchas de las zonas posibles que

RESERVAS COMPROBADAS
TOTAL: 658.000 MILLONES DE BARRILES

ARABIA SAUDITA	152
RESTO ORIENTE MEDIO	208
OTROS OPEP	90
NORTEAMERICA	40
OTROS NO-OPEP	65
AREA COMUNISTA	103

PRODUCCION ACUMULADA
TOTAL: 341.000 MILLONES DE BARRILES

ARABIA SAUDITA	23
RESTO ORIENTE MEDIO	61
OTROS OPEP	55
NORTEAMERICA	133
OTROS NO-OPEP	19
AREA COMUNISTA	50

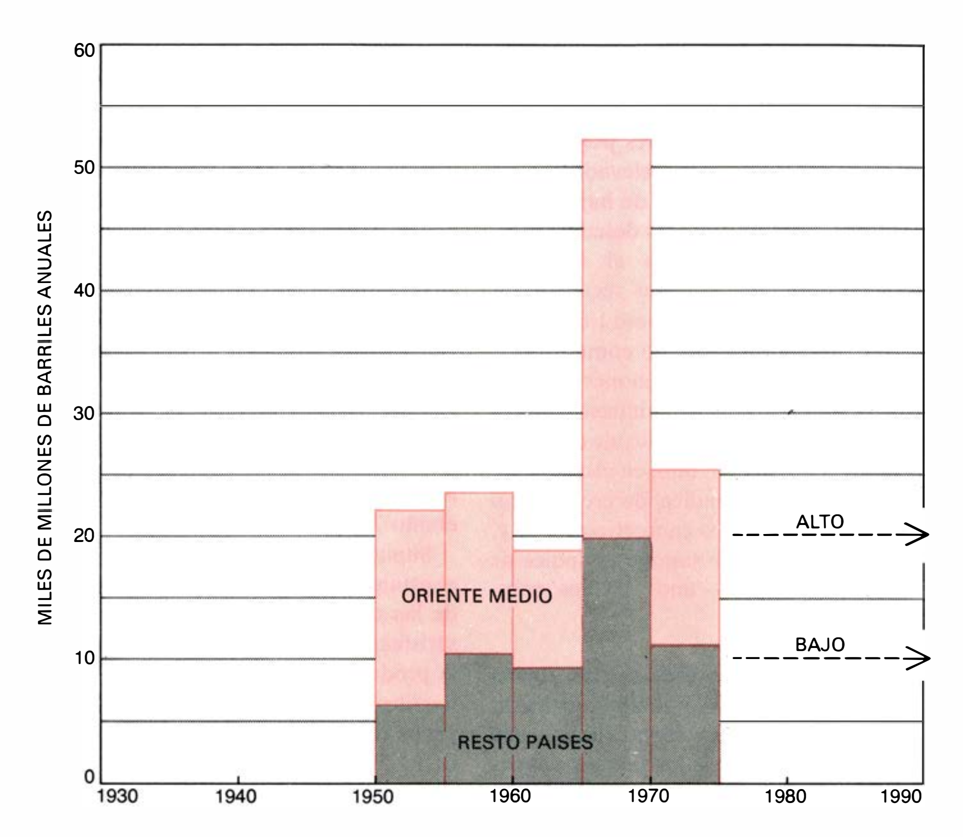
RESERVAS MUNDIALES DE PETROLEO existentes a finales de 1975, estimadas por el *Oil and Gas Journal (izquierda).* **De un total de 658.000 millones de barriles, 555.000 millones se encontraban fuera del mundo no-comunista. La mayoría de las reservas recuperables se encontraban en Oriente Medio, en tanto que la mayor producción tuvo lugar en Norteamérica** *(derecha).*

quedan en el mundo han sido ya evaluadas mediante pruebas sísmicas o sondeos exploratorios y no se han tenido pruebas de nuevos Orientes Medios. Probablemente se continuarán produciendo nuevos hallazgos en el propio Oriente Medio, pero el ritmo al cual se descubren las nuevas reservas ha empezado a descender.

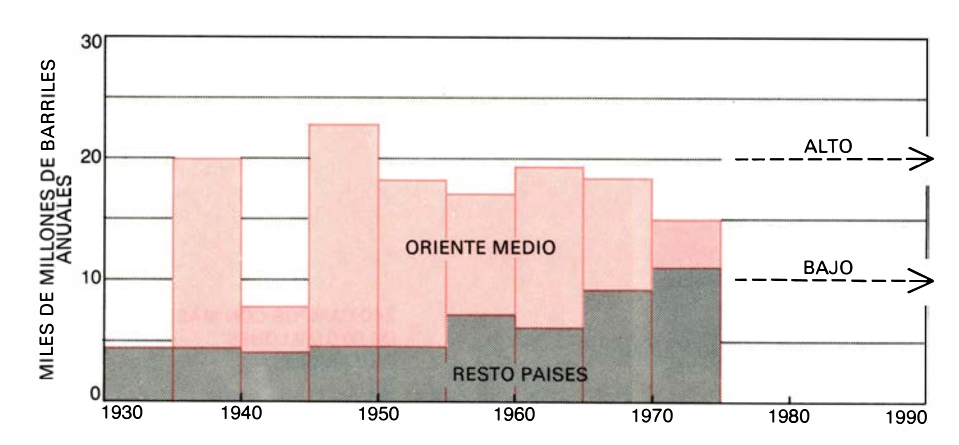
En Oriente Medio, y en otras zonas que se han identificado como las que gozan de mayor probabilidad en lo referente a contener amplias acumulaciones de petróleo, es en donde se efectuarán primero sondeos exploratorios. El índice de descubrimientos disminuirá cuando la investigación se traslade a zonas menos prometedoras. Al mismo tiempo, los incentivos para la exploración serán menores a medida que la probabilidad de los hallazgos de amplios campos descienda. Aunque se han descubierto unos 30.000 campos, alrededor del 75 por ciento de las reservas existentes se encuentran en un reducido número de ellos, cada uno de los cuales contiene más de 500 millones de barriles de reservas recuperables. Sólo 240 de tales campos se han descubierto en los últimos 100 años; y parece que no habrán muchos más. Hasta ahora, en ambientes hostiles tales como el mar del Norte, un campo con 500 millones de barriles recuperables es el mínimo que puede justificar los elevados costes de desarrollo. Con todo, parece improbable que se continúen haciendo nuevos descubrimientos a un ritmo de 15.000 millones de barriles por año, que ha sido el alcanzado en el último quinquenio.

¿Y qué sucederá con los aumentos de las reservas en campos descubiertos antes de 1975? La estimación de las reservas recuperables ha llegado a ser más precisa de lo que lo fue en los años 50, por lo que las correcciones positivas para los últimos 25 años no se van a repetir posiblemente en la misma proporción. En cambio, un origen más probable del incremento de las reservas reside en la mejora de las técnicas de recuperación. Hoy día, el índice medio mundial de recuperación es de un 30 por ciento; cada incremento del 1 por ciento en dicho índice aumentará, por consiguiente, las reservas recuperables en un trigésimo. Aplicando dicho factor a las reservas comprobadas de 555.000 millones de barriles y a la mitad de la producción acumulada de 291.000 millones de barriles (como una indicación de las reservas no propiamente consideradas recuperables) resulta una adición de alrededor de 25.000 millones de barriles por cada 1 por ciento de incremento en el índice de recuperación. John D. Moody y Robert E. Geiger, de la Mobil Oil Corporation, han sugerido que el índice mundial de recuperación puede eventualmente alcanzar el 40 por ciento, lo que significaría una adición de 250.000 millones de barriles a las reservas actuales en los campos conocidos.

Tomando todos estos factores en consideración, el Seminario basó sus proyectos en dos diferentes tasas de incremento para el período 1975-2000: 20.000 millones y 10.000 millones de barriles por año. La tasa superior supone que los nuevos descubrimientos descenderán gradualmente, y que las mejoras de la recuperación representarán una contribución significativa, respondiendo al menos de la mitad de las adiciones hasta el año 2000. La tasa inferior supone un descenso más rápido en el índice de nuevos descubrimientos y un rendimiento mucho menor



TASA DE DESCUBRIMIENTOS PETROLIFEROS en el pasado, representada como medias quinquenales de los incrementos anuales en reservas comprobadas, publicados por el *Oil and Gas Journal*. En este caso, tanto los nuevos descubrimientos como las nuevas evaluaciones de las reservas existentes se han asignado al año en que fueron realizados. Por ello, se destaca el período final de la década de los 60 en el que el mayor incremento en reservas conocidas resultó, principalmente, de las nuevas evaluaciones de los extensos campos petrolíferos ubicados en Oriente Medio.



LA TASA DE DESCUBRIMIENTOS muestra diferencias cuando sólo se toman en consideración, cada año, los nuevos descubrimientos realizados, introduciendo la revisión de las estimaciones del petróleo existente en los campos conocidos en el año en que fueron descubiertos. Este tratamiento tiende a nivelar las anteriores fluctuaciones quinquenales. Las líneas de trazos de la derecha indican supuestos altos o bajos de las futuras tasas de incremento de las reservas de petróleo.

de las mejoras en recuperación. Supusimos que entre los años 2000 y 2025 los incrementos disminuirían gradualmente a 4000 millones de barriles por año (en el caso de incrementos elevados) o de 3000 millones (en el caso de bajos incrementos), a medida que los descubrimientos totales se acercasen al inevitable techo: el total de reservas recuperables, que nosotros tomamos como 1,6 billones de barriles en las áreas no comunistas.

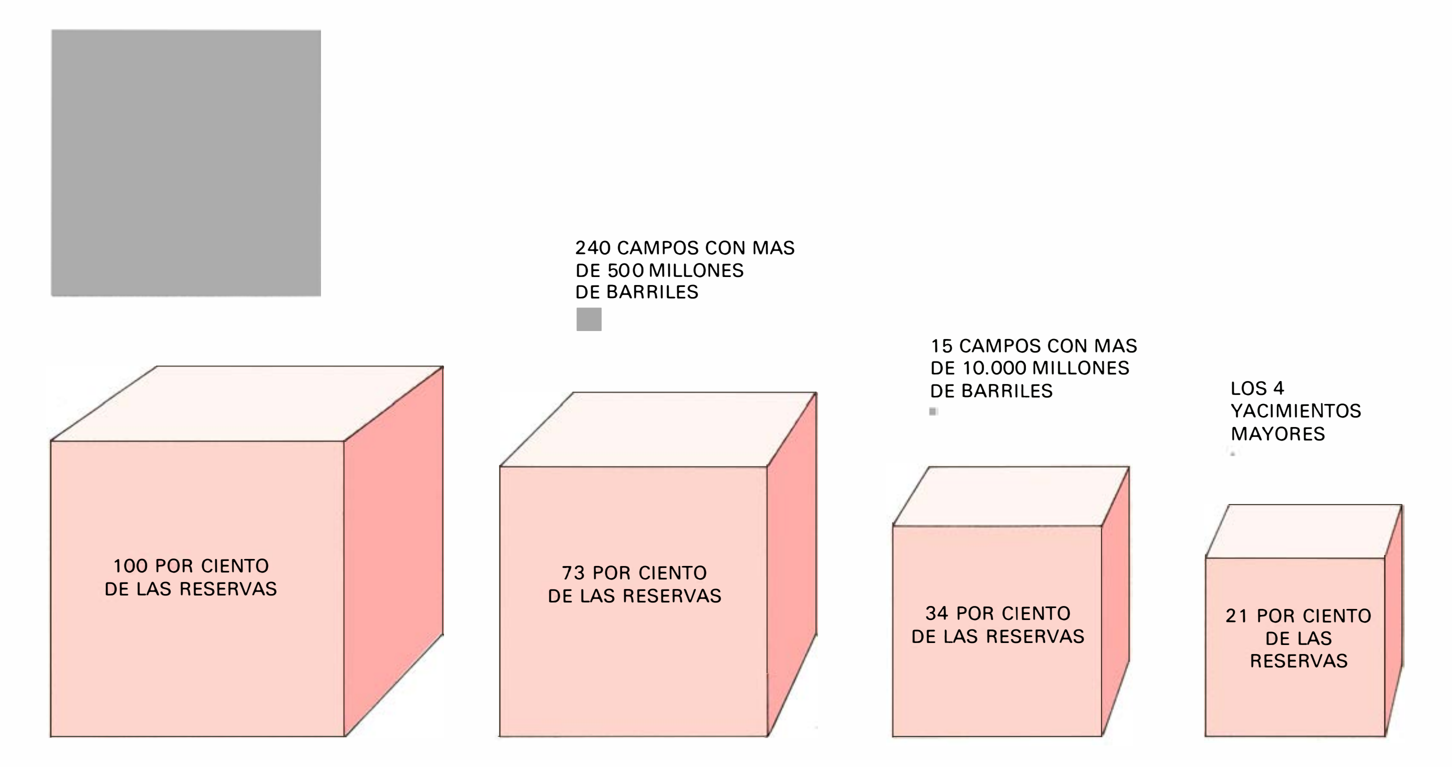
Estábamos ya en condiciones de desarrollar unos perfiles de suministro de petróleo para los dos modelos de demanda descritos más arriba: uno, en el que se supone un elevado índice de crecimiento económico y precios energéticos altos y, el otro, en el que se supone un índice de crecimiento bajo y unos precios constantes.

Supuesto un alto crecimiento y unos precios elevados –y también una tasa de incremento de las reservas elevada, que es lo más probable si los precios del petróleo se mantienen caros– la producción dejará de satisfacer la demanda en 1997. Esto es, aplicando la relación *R/P* de 15 a 1 a las reservas recuperables para 1997, se suministrará una producción potencial en ese año de 86 millones de barriles por día, mientras que la demanda, en el mismo año, será, de acuerdo con los cálculos, de 87 millones de barriles diarios. A partir de entonces, dado que las reservas continuarán descendiendo, la relación *R/P* de 15 a 1 reducirá la producción a 80 millones de barriles por día en el año 2000 y alrededor de 30 millones en el 2025. Si en este modelo sustituimos la tasa de bajo incremento de las reservas (que es menos probable, pero ciertamente posible), la producción dejará de satisfacer la demanda en 1990, cuando alcance sólo los 72 millones de barriles por día. Después descenderá a 50 millones de barriles por día en el año 2000 y alrededor de 20 millones en el año 2025.

Supuesto un bajo crecimiento, precios constantes y bajo índice de incremento de las reservas, la producción dejará de satisfacer la demanda en 1994, cuando la producción alcance los 66 millones de barriles por día, comparado con un proyecto de demanda, para el mismo año, de 66,5 millones de barriles por día. La producción caerá entonces a [illegible]3 millones de barriles por día y a unos 20 millones diarios en los años 2000 y 2025, respectivamente *(véase la ilustración inferior de la página siguiente)*. Si, ante un bajo crecimiento y precios constantes, la tasa de incremento de las reservas se mantiene alta, la producción podrá atender la demanda hasta el 2004, cuando se alcance una producción de 81 millones de barriles diarios.

El problema de los perfiles de suministro de crudos que acabo de describir es que parten de un supuesto algo irreal: que los países productores de petróleo continuarán incrementando su producción, para igualar la demanda, con sólo el límite de la relación *R/P* de 15 a 1. En realidad, algunos países productores no incluidos en la OPEP, para reducir su dependencia de la importación de crudos, pueden producir a una tasa superior a la indicada de 15 a 1. Y lo que es más importante, los países netamente exportadores (en particular, los miembros de la OPEP) pueden restringir su producción antes de alcanzar la relación de 15 a 1, con el propósito de conservar sus reservas. Es, por tanto, necesario modificar el análisis, considerando por separado la producción en los países miembros o no de la OPEP.

Para considerar separadamente la producción OPEP y no-OPEP habría que estimar la distribución de los futuros incrementos de las reservas entre ambos. Pero no se dispone de cálculos geológicos suficientemente establecidos. Nosotros supusimos que las nuevas aporta-

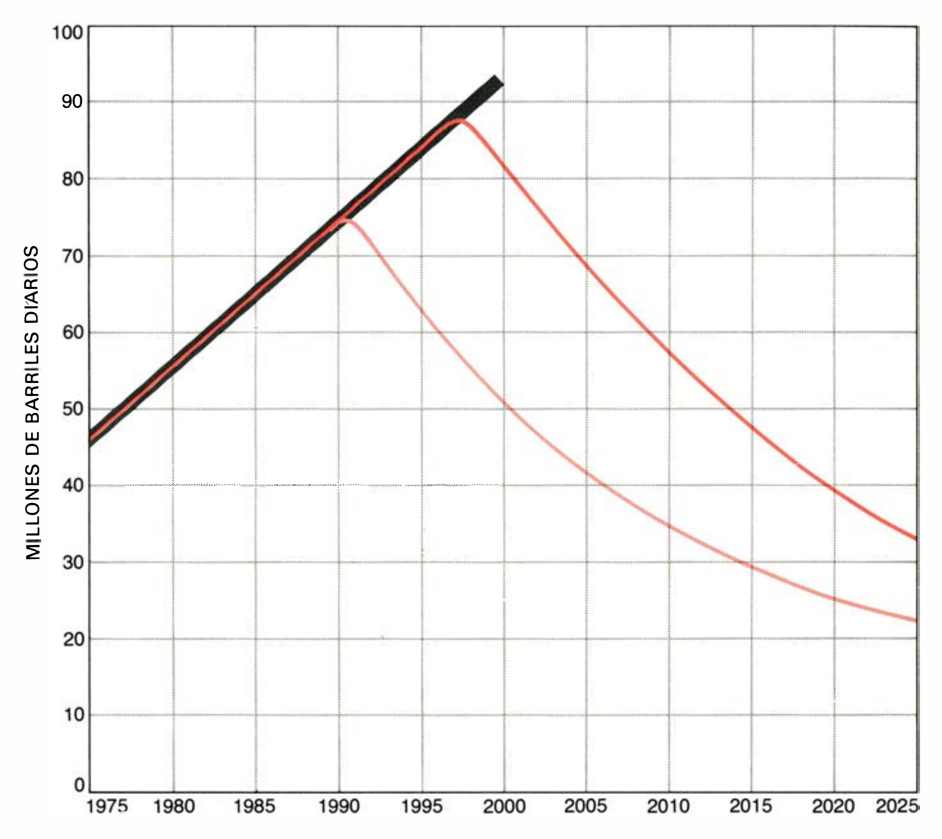


IMPORTANCIA de los grandes campos petrolíferos. De los 30.000 campos que se han descubierto, sólo 240 contienen más de 500 millones de barriles de reservas recuperables, si bien estos 240 aportan el 73 por ciento de las reservas conocidas en el mundo no-comunista. Sólo 15 campos contienen más de 10.000 millones de barriles, significando sus reservas más de un tercio de las totales. Los cuatro mayores campos petrolíferos conocidos, uno en Venezuela y los otros tres en Arabia, contienen juntos una quinta parte de las reservas mundiales existentes.

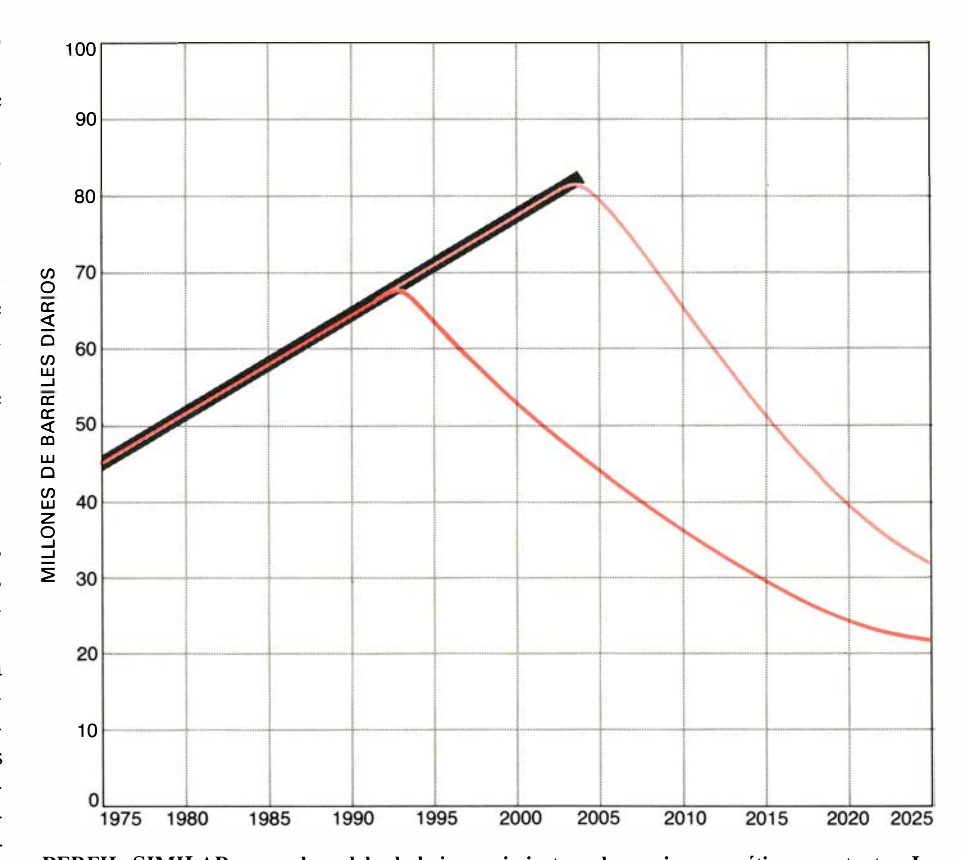
ciones se repartirían al 50 por ciento entre los dos grupos de países, es decir 10.000 millones y 5000 millones de barriles por año, en las hipótesis de altos y bajos índices del incremento de reservas, respectivamente. Muchas de las regiones no-OPEP siguen estando por explorar y deberían dar una mayor proporción de los nuevos descubrimientos; las mejoras en la recuperación, en cambio, serán más importantes dentro de la OPEP, donde están situadas el 80 por ciento de las reservas comprobadas.

Cada grupo nacional participante en el Seminario estimó la producción de su país para el período 1975-2000, y se hicieron cálculos similares para los otros países productores no pertenecientes a la OPEP. Para el período comprendido entre los años 2000 y 2025, la producción en los países no OPEP se estimó mediante los métodos referidos más arriba, con el nuevo supuesto de que la relación *R/P* tenderá a disminuir hasta 10 a 1 y que los incrementos de las reservas igualmente disminuirán *(véase la ilustración superior de la página siguiente).*

El cálculo de la producción OPEP en el futuro es algo más complicado de hacer, porque los distintos países tienen unas capacidades de producción, intereses económicos y sistemas de planificación muy diferentes. Lo que está claro es que, para situarse en el perfil de producción de petróleo más optimista, se requeriría que la producción OPEP alcanzara más de 50 millones de barriles por día, antes de 1995, lo que no será factible por razones tecnológicas y de planificación nacional. La "capacidad útil" corrientemente disponible en los países de la OPEP es de unos 37 millones de barriles por día, y en muchos países la expansión del equipamiento y la infraestructura de producción y transporte de crudos es improbable que tenga lugar en la proporción que sería necesaria. La influencia más significativa sobre la producción OPEP, sin embargo, será probablemente la insistencia de determinados países en prolongar la vida de sus propias reservas limitando la producción. Algunos gobiernos OPEP han sugerido ya límites específicos, que mantendrían la producción OPEP en 4,8 millones de barriles por día por debajo de su "capacidad útil". Lo más probable es que tales restricciones sean impuestas por los países que tienen una baja capacidad de asimilación de las rentas del petróleo (países como la Arabia Saudita, Kuwait o Libia, que obtienen unas ganancias superiores a las que pueden aplicar corriente-



PERFIL DE SUMINISTRO DE CRUDOS desarrollado aplicando una relación de reservas-producción *(R/P)* de 15 a 1 a las previsiones de reservas recuperables. Este es el perfil obtenido de acuerdo con el modelo de alto crecimiento y de precios energéticos elevados. Suponiendo un elevado índice de incremento de las reservas (que será la más probable si los precios fueran altos), la producción *(curva obscura)* dejará de satisfacer la demanda *(gris)* en 1997. A partir de un bajo índice de incremento la producción *(color claro)* dejará de satisfacer la demanda prevista en el año 1990.

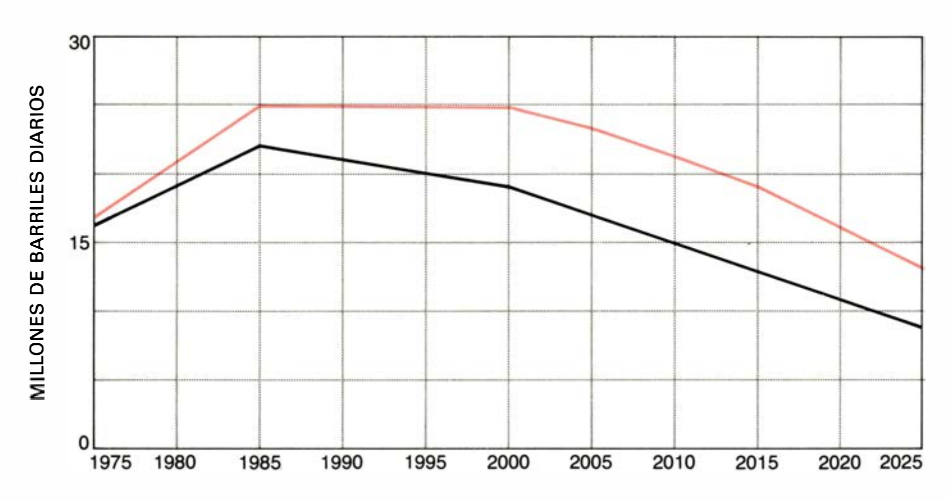


PERFIL SIMILAR para el modelo de bajo crecimiento y de precios energéticos constantes. La producción *(color obscuro)* empezará a desatender la demanda en 1994, considerando índices bajos de incremento de la reserva (situación probable). Si se suponen índices elevados, la producción *(color claro)* atenderá la demanda hasta el año 2004. No se presumen restricciones gubernamentales.

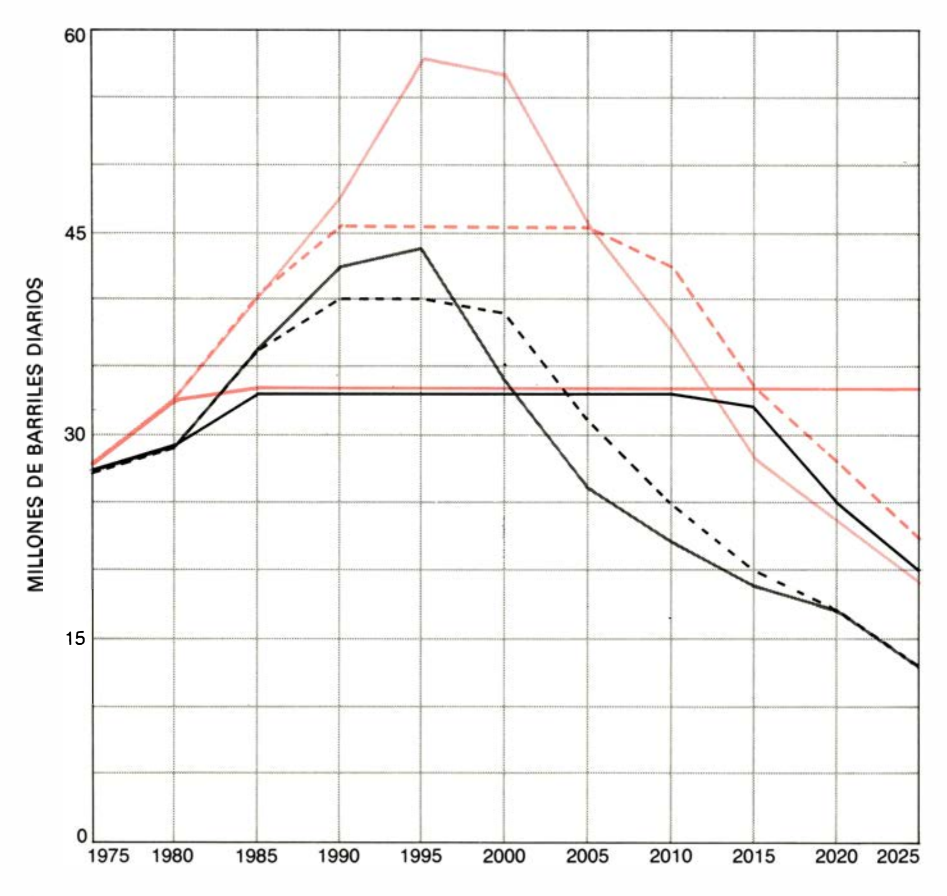
mente a sus necesidades domésticas). Por otra parte, Venezuela y Ecuador han alcanzado ya su techo de producción como para que puedan alargar mucho las rentas del petróleo, a pesar de que poseen una elevada capacidad de asimilación de tales beneficios.

Las restricciones impuestas por los gobiernos pueden tener una considerable incidencia en los ritmos potenciales de producción. Por ejemplo, de los países pertenecientes a la OPEP, solamente los de la Península Arábiga (Arabia Saudita, Kuwait, Qatar y los Emiratos Arabes) tienen el 58 por ciento de las reservas de la OPEP comprobadas. Si la relación R/P de 15 a 1 fuera la única restricción impuesta, podrían en la actualidad tener un ritmo de producción de 17.400 millones de barriles por año o 47 millones de barriles por día, que es casi igual al total de la producción no comunista. Pero ellos ya han situado su "capacidad útil" en 16,8 millones de barriles por día. Con todo, en 1975 produjeron sólo 11,3 millones de barriles por día y, si las restricciones que han sugerido llegan a imponerse, su producción futura podría mantenerse entonces en los 12,8 millones de barriles.

Teniendo en cuenta tales restricciones desarrollamos tres supuestos diferentes sobre los techos de producción en la Península Arábiga, techos cuyos efectos pudieran notarse antes de alcanzar la relación de R/P de 15 a 1. Si las limitaciones gubernamentales se fijaran en, o cerca de, los niveles actuales de producción, el suministro total de la Península Arábiga se mantendría alrededor de los 13 millones de barriles por día. Si Arabia Saudita permitiera un incremento de su producción hasta unos 15 millones de barriles por día (sobre los 9 millones, más o menos, obtenidos en 1977) la producción de la península se mantendría en 20 millones de barriles diarios. Y si Arabia Saudita elevara su producción hasta los 20 millones, el suministro total de la península se acercaría a los 25 millones de barriles por día.



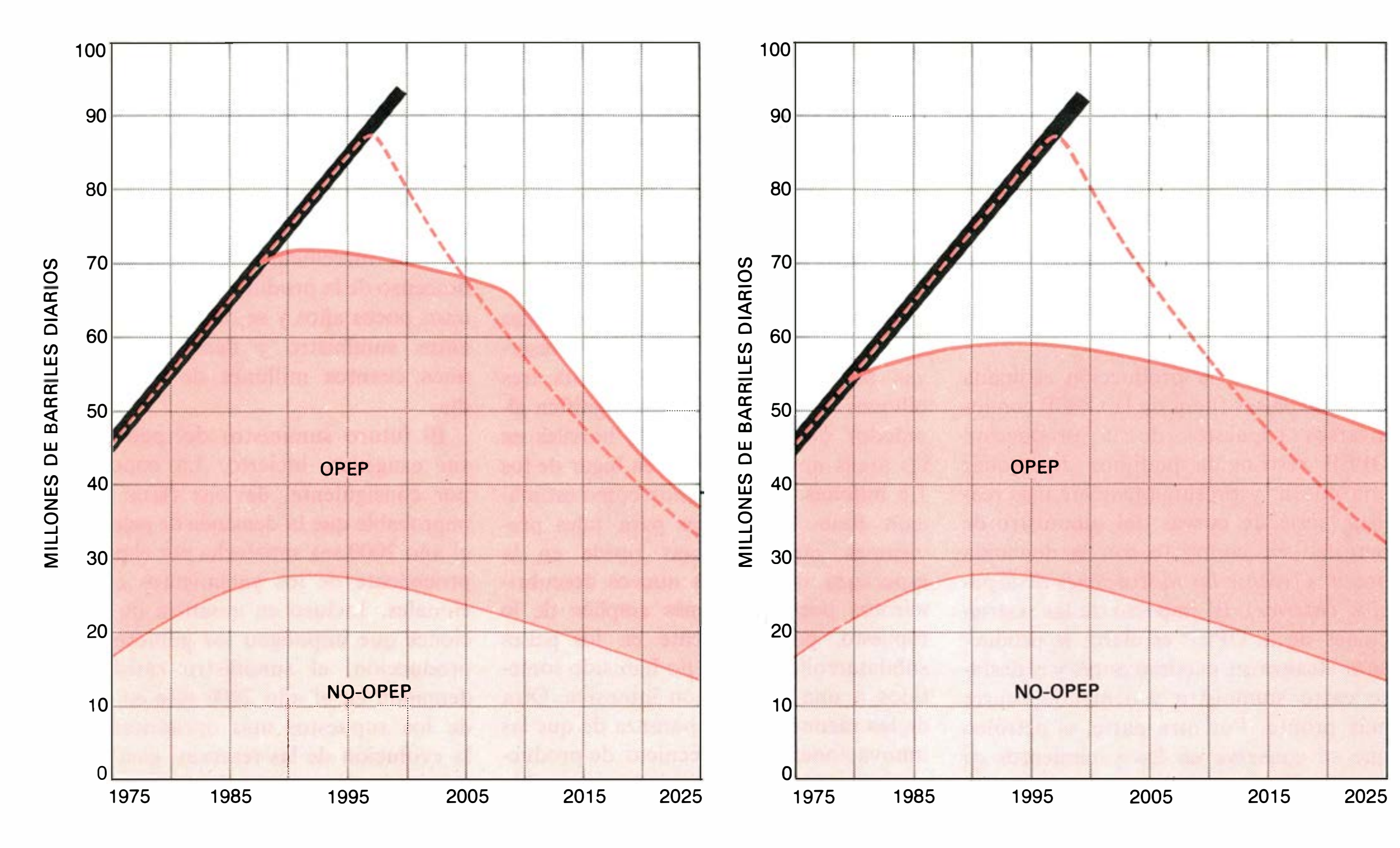
PRODUCCION NO-OPEP con vistas a considerar los efectos de las limitaciones gubernamentales. Las curvas muestran la producción no-OPEP estimada hasta el año 2000 y las previsiones hasta el 2025, suponiendo un declive en el incremento de las reservas y una disminución de la relación *R/P* para los modelos de alto crecimiento *(color)* y bajo crecimiento *(negro)*.



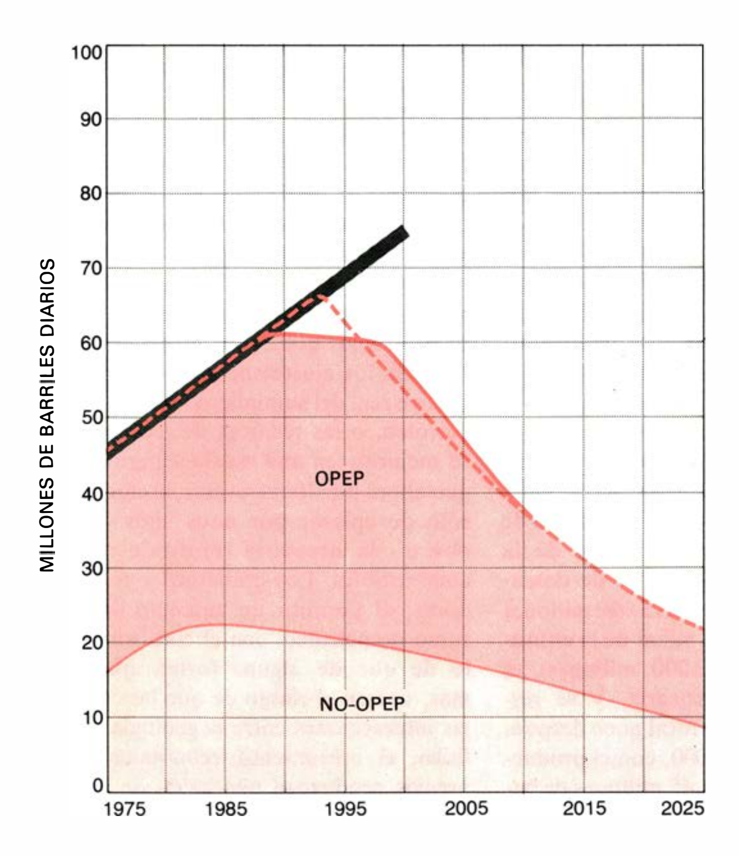
PREVISIONES DE LA PRODUCCION OPEP bajo varios supuestos de imposición de limitaciones por parte de los gobiernos. Para el modelo de alto crecimiento con elevados incrementos de las reservas se muestran tres curvas, correspondientes a restricciones gubernamentales de la producción a 33 millones de barriles por día *(línea continua coloreada)*, a 45 millones de barriles *(línea discontinua coloreada)*, o producción sólo restringida por la relación *R/P (línea de color claro)*. Para el modelo de bajo crecimiento y bajos incrementos de las reservas, la producción se mantiene en 33 millones de barriles por día *(línea negra)*, en 40 millones de barriles *(línea negra discontinua)* o sin restricciones *(gris)*. Las previsiones más allá del año 2000 resultan pura especulación.

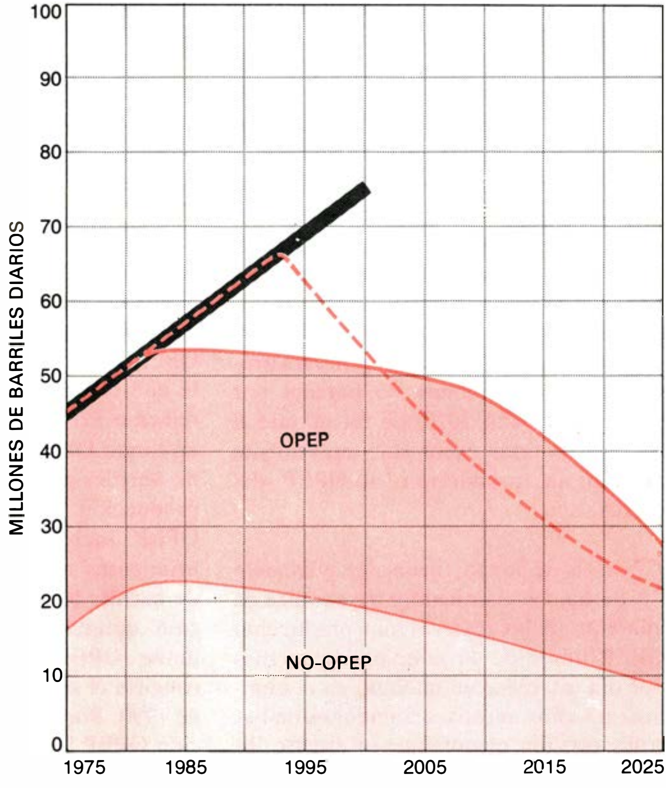
Los países de la OPEP no circunscritos a la Península Arábiga poseen una capacidad útil de 20,2 millones de barriles por día; las restricciones propuestas por tres de estos países pueden limitar la producción OPEP no árabe a 0,8 millones de barriles por día menos que dicha capacidad. Además, las reservas de algunos de estos países ya se están agotando con rapidez, de forma que las nuevas evaluaciones van a mantener probablemente la producción a su nivel actual más que a incrementarla de una manera significativa. Supusimos, por tanto, un límite tope de 20 millones de barriles por día para la producción OPEP, fuera de la Península Arábiga, límite que tiene en cuenta tanto las condiciones físicas impuestas por la disminución de las reservas como por las restricciones formales que pueden ser impuestas por los gobiernos.

La suma de los 20 millones de barriles por día de la producción no árabe con los tres supuestos techos de producción de la Península Arábiga (13, 20 y 25 millo-



PERFILES MAS REALISTAS teniendo en cuenta ciertas limitaciones gubernamentales. Las curvas corresponden al modelo de crecimiento elevado y de altos precios, así como un índice conspicuo del incremento de las reservas. Se muestra la producción no-OPEP, combinada con la producción OPEP, que está limitada a 45 millones de barriles por día *(izquierda)* y a 33 millones de barriles por día *(derecha)*, para dar la producción total en el mundo no-comunista *(línea de color fuerte)*. En ambos casos, la comparación con la situación no restrictiva *(línea discontinua)* muestra que la imposición de limitaciones gubernamentales situaría el suministro por debajo de la demanda *(gris)* antes, en 1981.





PERFIL REALISTA para el modelo de bajo crecimiento y de precio constante, así como de bajo índice del incremento de las reservas. Muestra efectos similares al anterior. Con la producción OPEP limitada a 40 millones de barriles por día *(izquierda)* o a 33 millones de barriles por día *(derecha)*, la producción deja de satisfacer la demanda antes que en el caso no restrictivo pero después la producción de las reservas que han sido conservadas mantiene el suministro notablemente constante por algún tiempo. (Las ilustraciones de este artículo se deben a Alan D. Iselin).

nes de barriles por día) da tres valores máximos de la producción OPEP potencial: 33, 40 y 45 millones de barriles por día. En el análisis revisado, se supone que la producción OPEP aumenta con la demanda sólo hasta alcanzar uno de estos tres límites, se mantiene a este nivel durante el tiempo en que la relación *R/P* sube por encima de 15 y, luego, disminuye al mismo tiempo que las reservas restantes.

Combinando la producción estimada para los países fuera de la OPEP con los diversos supuestos de la producción OPEP restringida pudimos desarrollar una nueva, y presumiblemente más realista, serie de curvas del suministro de petróleo en contraste con la demanda prevista *(véanse las ilustraciones de la página anterior).* El impacto de las restricciones de la OPEP es claro: la producción alcanza su máximo antes y el desfase entre suministro y demanda ocurre más pronto. Por otra parte, el petróleo que se conserve en los yacimientos en los años 90 ayudará a mantener las reservas y podrá ser explotado después del año 2000, haciendo el descenso de la producción algo menos precipitado.

Suponiendo el modelo de alto crecimiento y precios elevados y un conspicuo índice de aumento de las reservas, por ejemplo, el límite OPEP de 45 millones de barriles desplaza el máximo en el suministro de petróleo hasta 1989 (comparado con el de 1997 en el caso de ausencia de limitaciones) y hasta 71 millones de barriles por día (comparados con los 86 millones); pero el suministro podrá sostenerse en los 70 millones de barriles diarios, aproximadamente, durante casi 15 años. Con la producción OPEP limitada a 33 millones de barriles por día, en 1981 la demanda sobrepasará ya al suministro, a sólo 55 millones de barriles por día; la OPEP continuará produciendo 33 millones de barriles por día hasta el año 2025, de forma que el suministro total caerá sólo suavemente cuando la producción no-OPEP decrezca.

Con la demanda fijada por el modelo de bajo crecimiento y bajo índice de aumento de las reservas, una producción OPEP límite de 40 millones de barriles por día establece un máximo en el suministro 5 años antes y a 5 millones de barriles por día menos que en el caso ilimitado; pero de nuevo la curva del suministro se achata. Si la producción OPEP se limitara a 33 millones de barriles por día, el suministro total se restringiría a 55 millones en 1983 y caería sólo con el descenso de la producción OPEP a partir del año 2013, cuando la producción OPEP quedara finalmente por debajo de los 33 millones de barriles por día.

¿Son demasiado pesimistas estas perspectivas? Algunos observadores podrían considerar conservadores nuestros supuestos sobre incrementos de las reservas. Ha habido estimaciones optimistas que han fijado la totalidad de las reservas mundiales recuperables hasta tres billones de barriles, lo que significa alrededor de 2,4 billones de barriles en las áreas no comunistas, en lugar de los 1,6 billones que tomamos como estimación base. Una razón para tales previsiones tan optimistas reside en la esperanza de que los nuevos descubrimientos puedan ser más amplios de lo supuesto, particularmente en los países subdesarrollados que no han sido sometidos a una exploración intensiva. Otra de las razones es la esperanza de que las innovaciones en las técnicas de producción secundaria y terciaria harán posible recuperar cantidades de petróleo sustancialmente superiores a las supuestas en un principio. Ninguna de estas posibilidades puede descartarse. ¿Qué sucede entonces si se supone un incremento de las reservas comprobadas entre 1975 y el 2000 de 30.000 millones de barriles por año, en lugar de nuestra estimación más favorable de 20.000 millones?

Si no existieran restricciones en la producción OPEP, el índice de 30.000 millones permitiría abastecer la demanda potencial (en el caso del modelo de alto crecimiento) hasta los años 2005-2010, momento en el que la relación *R/P* alcanzaría el límite de 15 a 1 y la producción empezaría a declinar. Sin embargo, como he mostrado, el nivel de producción de petróleo al cual los países de la OPEP están dispuestos es más importante que el límite teórico impuesto por la relación *R/P.* Qué año podría alcanzarse un límite OPEP de, digamos, 45 millones de barriles por día, depende del nivel de producción de los países fuera de la OPEP. Incluso con un índice de descubrimientos no-OPEP de 15.000 millones de barriles por año (la mitad de la estimación optimista de 30.000 millones), el límite OPEP se alcanzaría, y se restringiría el suministro total poco después de 1990. Por el año 2000, con la producción OPEP limitada a 45 millones de barriles diarios, la producción total debería oscilar entre 75 y 80 millones.

Esto representaría un exceso de algo menos de 10 millones de barriles en la producción total, que se alcanzaría suponiendo un ritmo de descubrimientos de 20.000 millones de barriles por año, y aún se estaría alrededor de 16 millones de barriles diarios por debajo de la demanda implicada en el modelo de alto crecimiento. En otras palabras, adoptando incluso las estimaciones más optimistas para el incremento de las reservas, el descenso de la producción se retarda sólo unos pocos años y se reduce la distancia entre suministro y demanda en sólo unos cuantos millones de barriles por día.

El futuro suministro del petróleo es, sin exagerar, incierto. La conclusión, por consiguiente, deviene clara: resulta improbable que la demanda de petróleo en el año 2000 sea satisfecha por el petróleo procedente de los yacimientos convencionales. Incluso en ausencia de limitaciones que impongan los gobiernos a la producción, el suministro satisfará la demanda en el año 2000 sólo en el caso de los supuestos más optimistas sobre la evolución de las reservas. Una limitación más importante de la producción que el volumen actual de las mismas es probablemente el nivel de producción fijado por los países productores que tienen extensas reservas y una necesidad limitada de las rentas del petróleo. Los límites que probablemente fijarán tales miembros de la OPEP producirán el máximo de suministro, como más tarde, alrededor del año 1990; límites de producción más bajos anticiparán esta fecha a la década de los 80.

Aunque el final de la era del suministro creciente de petróleo está a lo sumo a una distancia de unos 15 años, puede ser seguida de otros 10 años de producción prácticamente constante, dando tiempo a los gobiernos y consumidores a hacer los ajustes necesarios antes del declive real del suministro. Si se descubre petróleo, o las técnicas de recuperación se mejoran, en una escala superior de lo que ahora puede entreverse, el efecto será sólo de aplazar por unos años –no de obviar– la necesaria transición a otros combustibles. Los gobiernos y consumidores, al permitir un aumento del consumo de petróleo, con el confiado anhelo de que de alguna forma aparecerá más, corren el riesgo de que las complejas interacciones entre la geología, la política, el crecimiento económico y los precios produzcan una caída de la producción aun antes de lo que habíamos pensado, incrementando entonces las dificultades de ajustarse a un mundo en el cual las existencias de petróleo son limitadas.

Leptones pesados

La clase de partículas elementales de la materia que incluyen el electrón y el muón tiene un nuevo miembro: la tau. Puede ser el primero de una secuencia de leptones pesados con carga

Martin L. Perl y William T. Kirk

La pequeña familia de partículas elementales conocida con el nombre de leptones (del griego "ligeros") puede distinguirse de las otras dos grandes clases de materia subatómica por varios criterios, de los cuales quizá la masa es el menos importante. Por ejemplo, los leptones se distinguen fundamentalmente de los hadrones, que suelen ser más pesados, por mostrarse insensibles a las fuerzas nucleares fuertes, las fuerzas dominantes de corto alcance que mantienen unidas las partículas del núcleo atómico. Los leptones comparten esta inmunidad ante las fuerzas fuertes con el fotón, el intermediador sin masa de las fuerzas electromagnéticas, que constituye una clase de la que es el único miembro. A diferencia del fotón, sin embargo, tanto los leptones como los hadrones son capaces de interaccionar mediante las fuerzas nucleares débiles, de alcance aun menor, que son las responsables de las desintegraciones radiactivas de las partículas nucleares.

A propósito de los leptones, una de las cosas más llamativas es su breve catálogo. Ello contrasta con lo que sucede con los hadrones, que han proliferado notablemente en los últimos años, y de los que hoy día existen varios centenares de partículas distintas, distribuidas en varias subclases. Los hadrones más familiares son los dos constituyentes principales del núcleo atómico, a saber: el protón y el neutrón; una de las últimas partículas que se ha añadido a esta clase es la psi o *J* descubierta en 1974. (Burton Richter, de la Universidad de Stanford, y Samuel C. C. Ting, del Instituto de Tecnología de Massachusetts, compartieron el premio Nobel de Física de 1976 por haber descubierto, independientemente, esta partícula.) Por otra parte, durante muchos años se ha creído que la familia de los leptones constaba únicamente de cuatro partículas (junto con las correspondientes antipartículas): el electrón, descubierto en la forma de rayos catódicos, hace más de 80 años; el muón, observado por primera vez en las cascadas de radiación cósmica hace unos 40 años; y los dos tipos de neutrinos, uno asociado con el electrón (llamado neutrino electrónico) y el otro asociado con el muón (el neutrino muónico). Los neutrinos, cuya existencia se sospechaba basándose en argumentos teóricos, fueron detectados por primera vez hace unos 20 años.

Recientemente se puso de manifiesto otro aspecto en que los leptones resultan especiales. Hay un cuerpo notable de pruebas que conducen a la conclusión de que los hadrones no son partículas elementales, sino que son estructuras compuestas de constituyentes simples que se denominan quarks. Con anterioridad al descubrimiento de la partícula psi, todos los hadrones conocidos podían explicarse a partir de la suposición de que representaran distintas combinaciones de tres clases diferentes de quarks [denominados "up" (hacia arriba)] "down" (hacia abajo) y "strange" (extraño)] y sus correspondientes antiquarks. El significado principal del descubrimiento de la partícula psi es el de suministrar una prueba muy vigorosa de la existencia de una cuarta clase de quarks, que había sido designada anteriormente como quark "encantado". De acuerdo con la situación actual del modelo quark, la partícula psi es un hadrón que consta de un quark con encanto y el correspondiente antiquark.

Carecemos de pruebas que indiquen que los leptones no sean objetos puntuales. Por tanto, parece que éstos, a diferencia de los hadrones, son realmente partículas elementales, en el sentido tradicional de ser indivisibles. La lista de las partículas conocidas que en este sentido pueden seguir considerándose elementales es corta: cuatro clases de quarks (y sus correspondientes antiquarks), cuatro clases de leptones (y los correspondientes antileptones) y el fotón (que coincide con su antipartícula). En este contexto, la búsqueda de nuevos miembros de la familia de los leptones ha adquirido nuevo interés, ya que tales partículas, si existieran, se las incluiría entre los pocos constituyentes genuinamente fundamentales de la materia.

Durante los últimos años, un grupo de físicos, dirigidos por uno de nosotros (Perl) ha estado realizando esta búsqueda en el Stanford Linear Accelerator Center (SLAC), como parte de un programa experimental más amplio llevado a cabo conjuntamente por varios grupos de investigación del SLAC y del Lawrence Berkeley Laboratory, de la Universidad de California. Teníamos a nuestra disposición una poderosa herramienta con la que, en principio, se podían crear nuevos leptones: el SPEAR, que es un anillo de almacenamiento electrón-positrón. En este aparato dos haces, que giran en direcciones contrarias, uno de materia (electrones) y otro de antimateria (positrones), se pueden hacer pasar el uno a través del otro, causándose ocasionalmente colisiones de alta energía en las que las partículas originales son aniquiladas y, del destello energético resultante, se crean otras nuevas. Con este sistema hemos encontrado pruebas de la existencia de una quinta clase de leptones. La nueva partícula, que tiene carga eléctrica (igual que el electrón y el muón), es mucho más pesada que los otros leptones ya conocidos; en realidad, pesa más que algunos hadrones. Vamos a explicar aquí la historia del descubrimiento del nuevo leptón pesado y de su antipartícula, que hemos llamado la tau y la antitau.

¿Cómo se hace para encontrar una nueva clase de partículas elementales? Ayudará el seguir unas cuantas reglas de conducta. En primer lugar, tener una idea clara de lo que estamos buscando; así sabremos cuándo hemos dado con ello. En segundo lugar, si el objeto debe producirse artificialmente. como sucede en este caso, encontrar



EL ANILLO de almacenamiento electrón-positrón del Stanford Linear Accelerator Center (SLAC) se ve desde la vertical de esta fotografía aérea de la zona experimental del SLAC. Los edificios están dispuestos aproximadamente en forma de abanico a partir del extremo de mayor energía del acelerador lineal de partículas de dos millas, localizado justamente fuera de la fotografía a la derecha. El anillo de almacenamiento para hacer chocar haces, denominado SPEAR, que se usó para crear el nuevo leptón pesado descrito en este artículo, es la estructura ovalada que aparece en la parte inferior. Los electrones y positrones que se generan en el acelerador lineal son inyectados en el SPEAR a través de dos brazos tangenciales. Los dos edificios, oblicuos a las secciones rectas del anillo, albergan las regiones de interacción; en éstas, los dos haces que giran en sentidos contrarios se hacen pasar el uno a través del otro, causándose colisiones materia-antimateria que conducen a la creación de nuevas partículas. El detector de partículas cargadas, cuyo esquema aparece más adelante, está en el mayor de los edificios de las zonas de interacción. El anillo de almacenamiento tiene unos 80 metros de diámetro. (La fotografía se reproduce con permiso del centro, el SLAC.)

PARTICULA	MASA	VIDA MEDIA	ANTIPARTICULA
ELECTRON (e^-)	0,51 MeV	ESTABLE	POSITRON (e^+)
NEUTRINO ELECTRONICO (ν_e)	0	ESTABLE	ANTINEUTRINO ELECTRONICO ($\bar{\nu}_e$)
MUON (μ^-)	106 MeV	$2,2 \times 10^{-6}$ SEGUNDOS	ANTIMUON (μ^+)
NEUTRINO MUONICO (ν_μ)	0	ESTABLE	ANTINEUTRINO MUONICO ($\bar{\nu}_\mu$)
TAU (τ^-)	1.800 A 1.900 MeV	MENOS DE 5×10^{-12} SEGUNDOS	ANTITAU (τ^+)
NEUTRINO TAUONICO (ν_τ)	MENOS DE 600 MeV (PUEDE SER 0)	NO SE SABE (PUEDE SER ESTABLE)	ANTINEUTRINO TAUONICO ($\bar{\nu}_\tau$)

LA FAMILIA DE LOS LEPTONES, se creía hasta hace poco, contenía sólo cuatro partículas que interaccionaban débilmente (y sus correspondientes antipartículas): el electrón, el muón y dos clases de neutrinos. La masa del leptón pesado y cargado que hemos descubierto, la tau, y su antipartícula, la antitau (dada aquí en MeV, o millones de electrón-voltios), es aproximadamente 4000 veces la masa del electrón y 20 veces la del muón. No se conoce la razón de estas relaciones entre masas. Los muones se desintegran en electrones mediante las interacciones débiles. Las tau se desintegran aún más rápidamente en electrones, muones u otras partículas, también mediante las interacciones débiles. No se han observado todos los modos de desintegración de la tau y la antitau.

una forma de hacerlo en grandes cantidades. En tercer lugar, asegurarse de que el nuevo objeto tenga alguna característica peculiar que nos descubra en qué momento hemos dado con él.

Para hacernos una idea de qué era lo que andábamos buscando tuvimos en cuenta las propiedades del electrón y del muón. En otras palabras, decidimos buscar una partícula que poseyera carga eléctrica −1 o +1 y que sufriera las fuerzas electromagnéticas y débiles, pero no las fuerzas fuertes. En ese instante se plantean dos preguntas. Primero, en nuestra búsqueda del nuevo leptón cargado, ¿qué intervalo de masas debemos considerar? Segundo, teniendo en cuenta que el electrón está dotado de una duración indefinida y que el muón se desintegra en unas dos millonésimas de segundo, ¿cuál hemos de esperar que sea la vida media de la nueva partícula?

El problema de la masa era complicado, pues no había (y aún no hay) una teoría que explique las masas observadas del muón y del electrón ni la razón entre ambas (aproximadamente 200 a 1). Todo lo que sabíamos hace cuatro años es que algunos experimentos llevados a cabo en el ADONE, una máquina italiana en la que chocan haces de electrones y positrones, habían establecido una cota inferior de alrededor de 1000 millones de electrón-voltios (MeV) sobre la masa de cualquier nuevo leptón cargado: una energía equivalente aproximadamente a 10 veces la masa del muón y 2000 veces la masa del electrón. Nuestro grupo del SLAC no tenía ninguna noción sobre si la masa del nuevo leptón cargado, suponiendo que existiera, estaría dentro de los límites de nuestro equipo experimental, que podía detectar partículas con una masa de hasta 3500 MeV.

La pregunta sobre la vida media de la hipotética partícula resultó tener una mejor contestación teórica: menor de una cien mil millonésima de segundo para todo leptón cargado cuya masa fuera mayor de 1000 MeV. El leptón cargado más ligero, el electrón, es estable por la simple razón de que no hay una partícula cargada más ligera en la que pueda desintegrarse. El muón es inestable porque puede decaer en un electrón. La desintegración de muón a electrón no tiene lugar, sin embargo, de la forma más sencilla que podamos concebir, que sería que el muón se cambiara espontáneamente en un electrón y un fotón mediante una interacción electromagnética, aun cuando en este proceso se liberaría energía más que suficiente para producir un fotón. En lugar de esto se observa que la desintegración tiene lugar a través de una interacción débil y se producen tres partículas: un electrón, un antineutrino electrónico y un neutrino muónico.

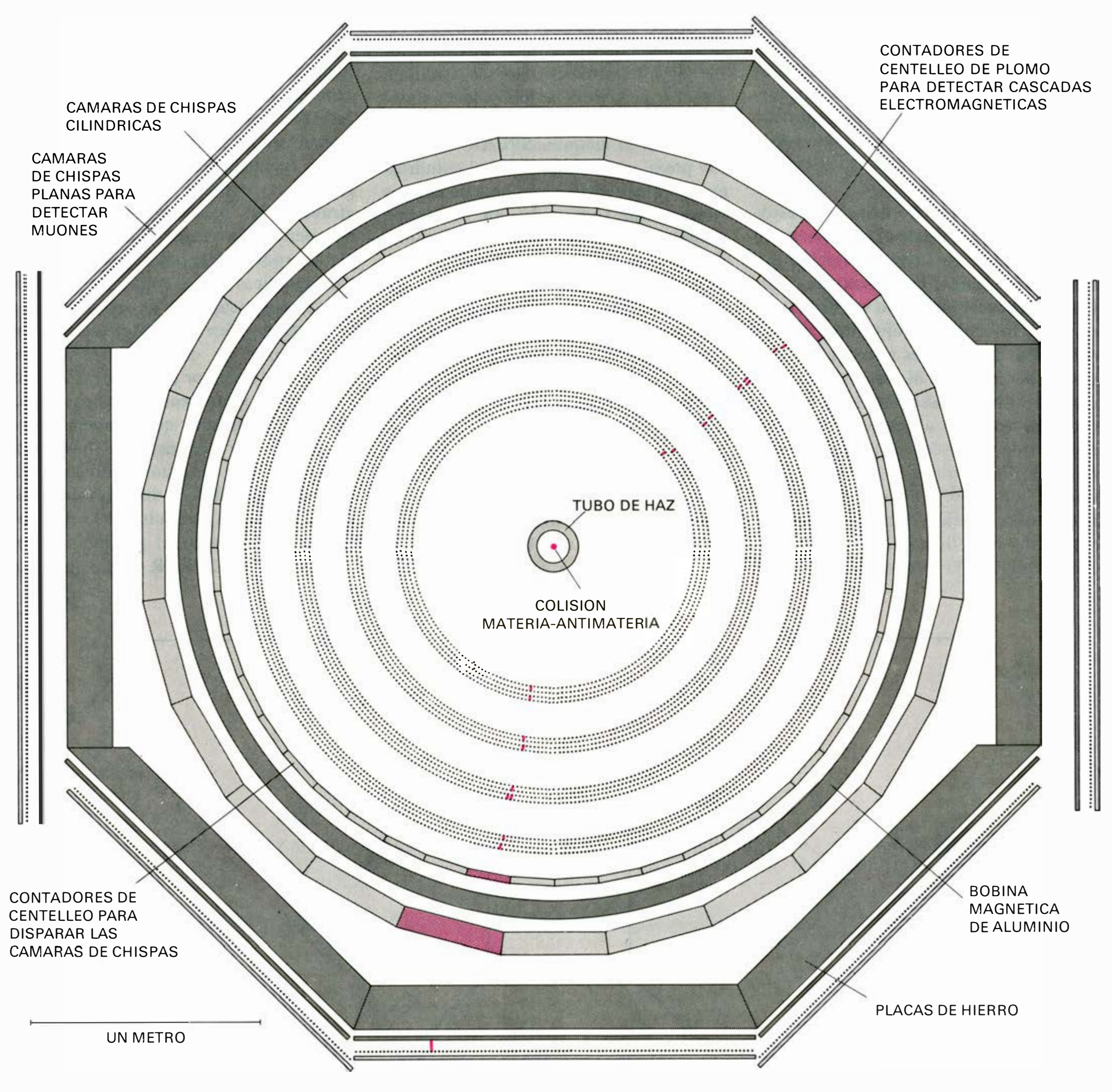
Los físicos explican este extraño comportamiento invocando una regla empírica cuyo significado básico no ha sido aclarado todavía. Asignan al electrón y al muón propiedades intrínsecas separadas, que denominaremos aquí "carácter electrónico" y "carácter muónico", y postulan que cada una de estas propiedades aparece en un grupo de cuatro partículas que están relacionadas. Así el electrón, el neutrino electrónico, el antielectrón (o positrón) y el antineutrino electrónico tienen la propiedad de carácter electrónico (o carácter antielectrónico en el caso de las antipartículas), mientras que el muón, el neutrino muónico, el antimuón y el antineutrino muónico exhiben todos carácter muónico (o carácter antimuónico). La regla establece simplemente que, en cualquier proceso de interacción entre partículas o de desintegración de éstas en el que intervengan leptones, las propiedades de carácter electrónico y de carácter muónico deben conservarse por separado.

Síguese, pues, que la simple desintegración del muón en un electrón y un fotón no es posible, ya que, en el proceso, el carácter muónico del muón debería cambiar en el carácter electrónico del electrón. La desintegración débil del muón en un electrón, un antineutrino electrónico y un neutrino muónico (que es más compleja) sí ocurre, y ello porque el neutrino muónico conserva el carácter muónico del muón mientras que el carácter electrónico del electrón se cancela exactamente con el carácter antielectrónico del antineutrino electrónico. (En la práctica, los físicos de partículas emplean los términos número leptónico electrónico y número leptónico muónico para indicar las propiedades de carácter electrónico y carácter muónico; la regla en cuestión se denomina de la conservación del número leptónico. Nuestro punto de vista es, sin embargo, que estos términos formales tienden a obscurecer el misterio fundamental de estas propiedades separadas, sin igual, del electrón y del muón.)

En cualquier caso teníamos que decidirnos por alguna propiedad comparable de este hipotético leptón pesado, dotado de carga, que buscábamos. Una posibilidad era asignarle un número leptónico electrónico (o carácter electrónico) a este nuevo leptón, permitiéndole por tanto desintegrarse electromagnéticamente en un electrón y un fotón. Otra posibilidad, más intrigante, era suponer que el nuevo par leptón-antileptón tenía su propio número leptónico distinto y su par asociado neutrino-antineutrino. De acuerdo con este punto de vista, el electrón y el muón podrían ser sólo el comienzo de una secuencia de leptones cargados, cada uno de ellos con un "carácter" distinto. El nombre que se adoptó entonces para estas

partículas altamente especulativas fue el de leptones pesados cargados secuenciales; el adjetivo "pesado" indicaba que sus masas deberían ser mayores que las del electrón y del muón. En cuanto obtuvimos pruebas fragmentarias de haber encontrado tal partícula empezamos a usar el símbolo U [(inicial de la palabra inglesa "unknown" (desconocida)], pero ahora que tenemos pruebas substanciales de su existencia la hemos llamado tau, que es la primera letra de la palabra griega *τριτον*, que significa "tercero". El nombre denota que tau es el tercer leptón cargado en una secuencia que empieza con el electrón y el muón.

Como hemos dicho antes, la búsqueda efectiva de una nueva partícula elemental requiere tener alguna idea previa de sus propiedades, un método para producir la partícula en cantidades sufi-



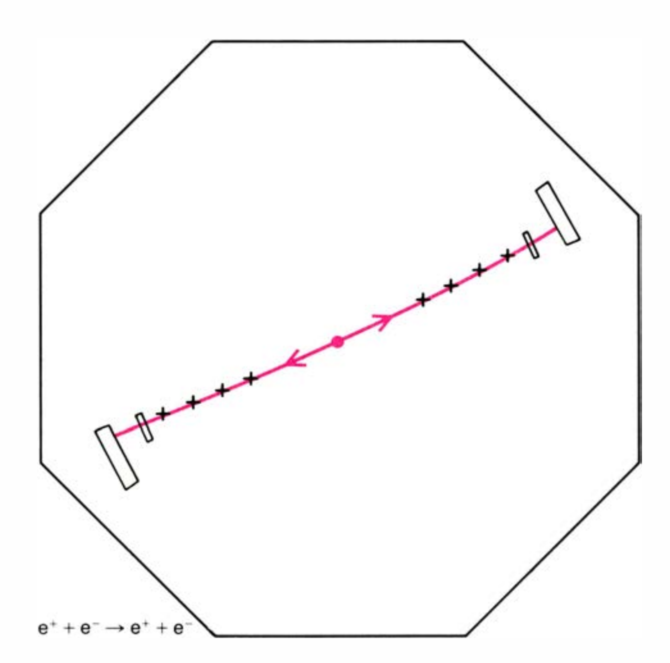
DETECTOR de partículas polivalente. Rodea una región de interacción del anillo de almacenamiento electrón-positrón llamado SPEAR; aparece en una sección recta, con las trayectorias de los haces que chocan perpendiculares al plano del diagrama y el punto de colisión en el plano del diagrama en el centro *(punto coloreado)*. La parte clíndrica del interior del detector contiene cuatro cámaras de chispas concéntricas, cada una de las cuales consta de cuatro capas de alambres que forman un apretado enrejado *(puntos negros)*. En el aparato hay unos 100.000 alambres, siendo la distancia entre ellos aproximadamente de 1 milímetro. Cuando el detector está en marcha, se aplica un alto voltaje entre pares de capas adyacentes de alambres, y saltan chispas *(rayas coloreadas)* que marcan las trazas de ionización creadas por las partículas cargadas cuando pasan a través de los gases inertes que llenan las cámaras. Exteriormente a las cámaras de chispas, hay dos sistemas de contadores de partículas cargadas, separados por una bobina magnética de aluminio. Cuando el sistema interior de controles de centelleo detecta el paso de dos o más partículas cargadas, las cámaras de chispas cilíndricas se disparan y se registran los caminos de las partículas. (Los caminos de las partículas son ligeramente curvados por un campo magnético; la curvatura indica el momento de la partícula). El sistema exterior de contadores de centelleo de plomo registra las características cascadas electromagnéticas, generadas por los distintos tipos de partículas. Encerrando todo este dispositivo existe un conjunto octogonal de gruesas placas de hierro. De todas las partículas que pueden producirse en una colisión electrón-positrón, las que penetran a través del hierro y se detectan en el sistema exterior de cámaras de chispas planas son los muones.

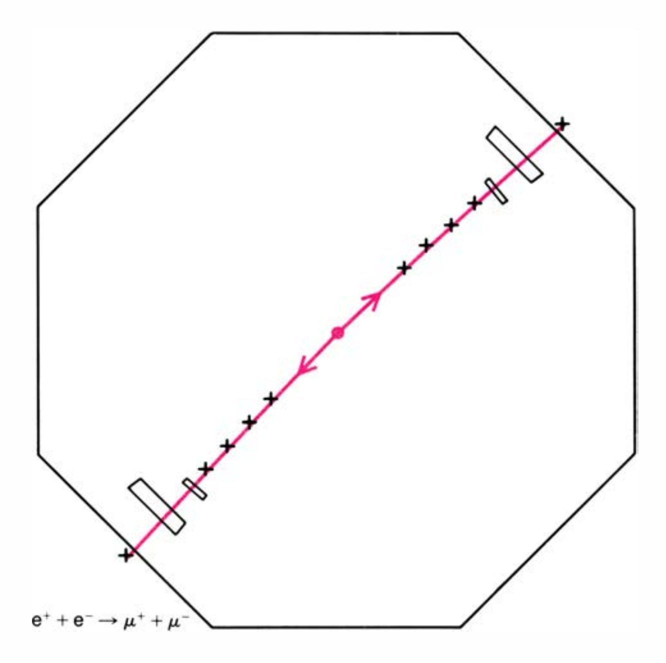
cientes y un método para distinguir la nueva partícula de las otras ya conocidas. Los experimentos que utilizan el anillo de almacenamiento electrón-positrón, llamado SPEAR, en el SLAC pueden cumplir las dos últimas condiciones. Esta máquina, que se puso en funcionamiento en 1972, consta de unos 100 imanes colocados formando un anillo de unos 80 metros de diámetro. Para el SPEAR, los haces electrón y positrón, que son generados en el acelerador lineal de dos millas del SLAC, se inyectan en el anillo de almacenamiento durante un "tiempo de llenado" que puede durar entre 10 y 30 minutos. Los haces circulan en una cámara de vacío que pasa a través del anillo de imanes; éstos desvían y enfocan los haces de forma que los mantienen en órbitas estables durante varias horas. Los haces que circulan por el acelerador sólo se encuentran en dos puntos diametralmente opuestos del anillo, donde se les hace pasar a través de dos regiones de interacción. Para maximizar la probabilidad de colisiones, cada haz contiene alrededor de 100.000 millones de partículas en un único "racimo" muy denso que sólo tiene unos centímetros de largo. Aunque cada racimo hace una revolución completa alrededor del anillo más de un millón de veces cada segundo, la probabilidad de que una partícula de un haz entre en colisión directa con otra partícula del otro haz es tan remota que una colisión sucede, en general, sólo una vez cada pocos segundos.

Para estudiar todas las partículas que puedan producirse en una aniquilación electrón-positrón, dos grupos de físicos del SLAC unieron sus esfuerzos, hace varios años, con dos grupos del Lawrence Berkeley Laboratory para construir un detector de partículas polivalente al objeto de rodear una de las regiones de interacción de la máquina SPEAR. El detector posee una sección central cilíndrica que contiene cuatro cámaras de chispas concéntricas sumergidas en un campo magnético intenso. Rodeando las cámaras de chispas hay un sistema de contadores de centelleo para detectar partículas cargadas. Cuando los contadores detectan dos o más partículas cargadas, se disparan las cámaras de chispas internas y se registran los caminos de las partículas cargadas mediante la ayuda de un ordenador o una cinta magnética. El ordenador puede reconstruir entonces los caminos de las partículas, dando una "imagen" diagramática del suceso.

En la parte exterior de los contadores de centelleo hay una bobina magnética de aluminio, y esto va seguido hacia la parte exterior por otro sistema cilíndrico de contadores para detectar cascadas electromagnéticas. Estos contadores permiten distinguir electrones o positrones, que generan grandes cascadas, de hadrones o muones, que producen cascadas menores. En la parte exterior de todos estos instrumentos hay un conjunto octogonal de placas de hierro de, al menos, 8 pulgadas de espesor (1 pulgada equivale a 25,4 milímetros) y a continuación más cámaras de chispas. En general, los hadrones no pueden atravesar las placas de hierro porque interaccionan con los núcleos de hierro mediante las fuerzas fuertes. Los electrones y positrones también son incapaces de atravesar las placas de hierro, ya que han perdido la mayor parte de su energía al producir las grandes cascadas electromagnéticas. Por el contrario, los muones sí son capaces de atravesar las planchas de hierro, y son detectados en las cámaras de chispas externas. De aquí que el detector puede no sólo medir la dirección y momento de las partículas recién creadas sino que también puede identificar por separado los hadrones, los electrones y los muones. Precisamente, esta posibilidad de distinguir las diferentes clases de partículas es lo que nos ha permitido hallar el par tau-antitau.

El anillo del SPEAR puede ajustarse





CUATRO DIAGRAMAS SIMPLIFICADOS de posibles resultados de una aniquilación mutua de un electrón y un positrón. Para su dibujo, se ha echado mano de una pantalla ayudada por un ordenador conectado al detector de partículas cargadas de la máquina de colisión de haces de Stanford. En representaciones de esta clase, las descargas eléctricas causadas por el paso de partículas cargadas a través de las cámaras de chispas cilíndricas interiores o de las planas exteriores se simbolizan por pequeñas cruces negras. Solamente se muestran aquellos contadores que han detectado el paso de partículas; los rectángulos pequeños son los contadores de disparo internos, y los rectángulos mayores, y más planos, son los contadores más externos de cascadas electromagnéticas. El esquema octogonal indica la superficie exterior de las placas de hierro. Las líneas coloreadas que conectan las señales de detección están curvadas en arcos de círculo debido al fuerte campo magnético existente dentro de la bobina magnética, pero son rectas fuera de dicha bobina. El dibujo de la izquierda muestra un típico suceso electrón-positrón, en el cual el par original electrón-positrón desaparece, y, de la energía resultante, un nuevo par electrón-positrón emerge; las dos partículas salen en direcciones opuestas llevándose cada una de ellas la mitad del total de la energía de las partículas que chocaron. (Al salir

para que almacene haces de electrones y positrones de una energía de hasta 4000 millones de electrón-voltios (GeV) en cada haz. Puesto que las colisiones ocurren entre materia y antimateria, un posible resultado de esta colisión es lo que se denomina una reacción de aniquilación. La reacción tiene realmente lugar en dos pasos. En primer lugar, el electrón y positrón que chocan desaparecen; al hacer esto, crean un estado de vida muy corta consistente en energía electromagnética pura y denominada fotón virtual. Entonces, después de un tiempo extraordinariamente breve (10^{-25} segundos), el fotón virtual se materializa en una cualquiera de un número muy grande de posibles combinaciones de nuevas partículas. Entre los posibles resultados se cuentan los siguientes: que se vuelve a crear un par electrón-positrón, la creación de un par muón-antimuón, la creación de un par hadrón-antihadrón (por ejemplo, un protón y un antiprotón) o la creación de un gran número de hadrones. Nosotros nos habíamos propuesto emplear el mismo método para producir un par leptón-antileptón pesado.

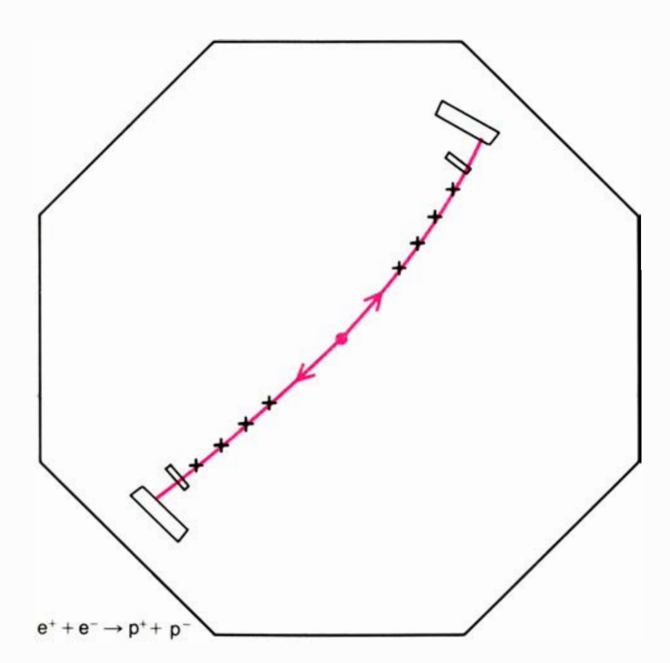
Si tales pares de nuevos leptones pesados eran realmente creados en las colisiones electrón-positrón en el SPEAR, ¿cómo seríamos capaces de reconocerlos? Nuestro cálculo anterior de la vida media de la tau –menor de una 100.000 millonésima de segundo– significaba que una tau recién creada viajaría menos de un centímetro desde el punto de origen al punto de desintegración, aun suponiendo que su velocidad fuera aproximadamente igual a la de la luz (30.000 millones de centímetros por segundo). Este camino es demasiado corto para que pueda detectarse de forma directa por nuestro dispositivo experimental; por tanto, debíamos buscar una forma de reconocer las partículas tau indirectamente, identificando alguna estructura característica que apareciera en el momento de su desintegración.

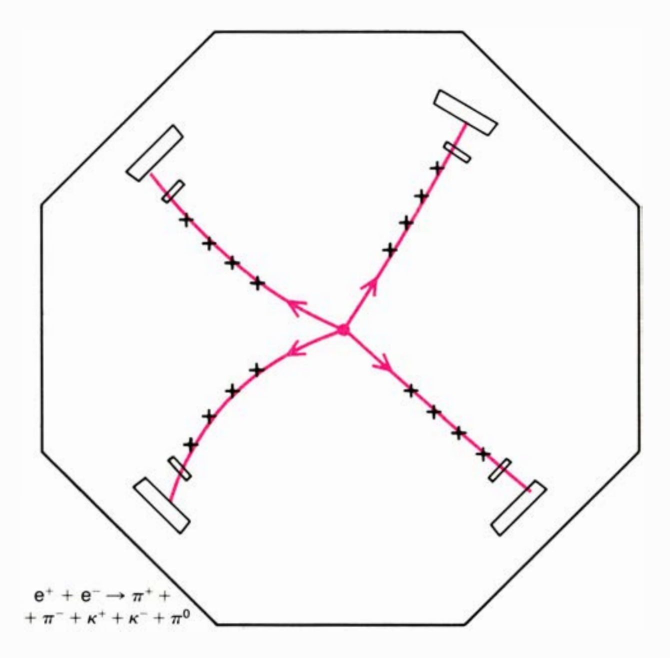
Nuestra decisión de atribuir un nuevo número leptónico, o "carácter tauónico", a la tau y su neutrino, significaba que nosotros no esperábamos que la tau se desintegrara electromagnéticamente en un electrón o un muón acompañados por la emisión de un fotón. Por el contrario, supusimos que la tau, igual que el muón, se desintegraría mediante las interacciones débiles, y que serían posibles varios modos distintos de desintegración debido a su gran masa. Seleccionamos dos de estos posibles modos de desintegrarse débilmente, para centrar en ellos nuestra atención.

En el primer modo de desintegración la tau se desintegra en un neutrino de tipo tau, un muón y un antineutrino muónico, y, en el proceso inverso, la antitau se desintegra en un antineutrino de tipo tau, un antimuón y un neutrino muónico. En el segundo modo de desintegrarse una tau decae en un neutrino de tipo tau, un electrón y un antineutrino electrónico, y la antitau se desintegra en un antineutrino de tipo tau, un positrón y un neutrino electrónico.

Se eligieron estos dos modos de desintegración porque esperábamos que ocurrieran frecuentemente. Aun más importante, estos modos de desintegración que parecen ser muy complicados resultan, en realidad, muy simples cuando se analizan en el contexto de nuestro experimento. La razón es que todos los neutrinos y antineutrinos producidos en la desintegración son partículas tan evasivas que el aparato experimental no las "ve" en absoluto. Entonces, lo que el aparato registra en realidad, en cada caso, es únicamente la traza de dos partículas cargadas: bien sea un electrón y un antimuón, bien sea un positrón y un muón.

El producto final tiene una "marca" experimental característica por dos razones. En primer lugar, la aparición de un electrón y un muón (o sus antipartículas)





del punto de choque, las partículas cargadas negativamente se curvan hacia la derecha y las partículas cargadas positivamente se curvan hacia la izquierda, perpendicularmente al campo magnético; por tanto, la traza de la partícula en la posición de las ocho en punto es la del electrón, mientras que la de las dos en punto es la del positrón.) El segundo diagrama muestra un suceso muón negativo (en la posición de las cuatro en punto) y un antimuón positivo (en la posición de las diez). A diferencia del electrón y del positrón, que pierden la mayor parte de su energía al generar las cascadas electromagnéticas y, por consiguiente, no pueden penetrar en las placas de hierro, los muones atraviesan el hierro y son detectados por las cámaras de chispas más externas. En el tercer diagrama, el electrón y el positrón originales se aniquilan mutuamente para formar un par de hadrones, en este caso un protón y un antiprotón. En el cuarto diagrama, las trazas de cuatro hadrones son observadas en el detector: dos piones cargados y dos kaones cargados. Una quinta partícula, un pión neutro, es también producida en la reacción de aniquilación, pero como no tiene carga no puede ser directamente detectada. Ninguno de los hadrones es capaz de atravesar las placas de hierro. Las fórmulas que acompañan a cada una de las distintas ilustraciones descubren las interacciones en la notación abreviada que usan los físicos de partículas.

en el estado final parece violar el principio de la conservación del número leptónico; sin embargo, esto no es cierto, ya que los neutrinos no detectados restablecen el balance. En segundo lugar, falta mucha energía, pero esto también se explica mediante los neutrinos que no son vistos.

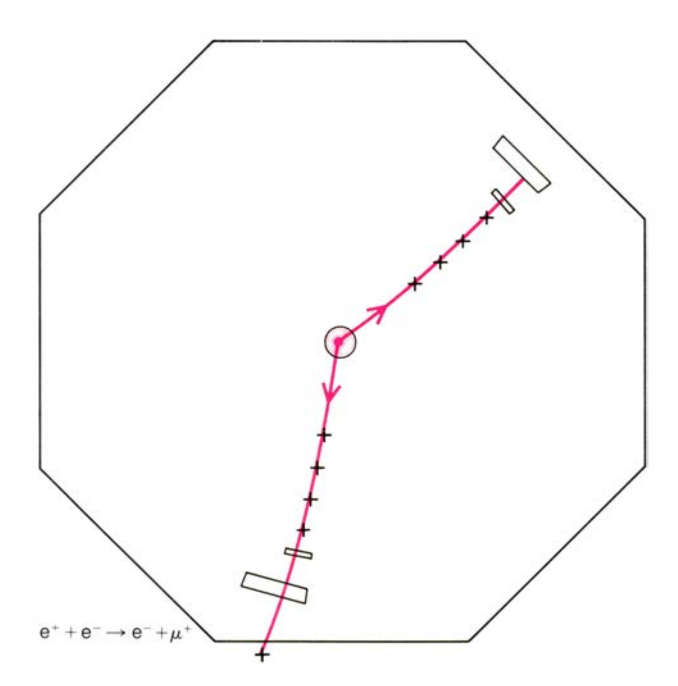
Encontramos las primeras pruebas de estos sucesos electrón-muón en 1974. De una muestra de 10.000 sucesos de todos los tipos, identificamos 24 como sucesos electrón-muón. Aunque el número de sucesos era bajo, nos sentimos animados, pues si estos sucesos eran reales, significaba que la masa de la tau estaba dentro del intervalo de energías del sistema SPEAR.

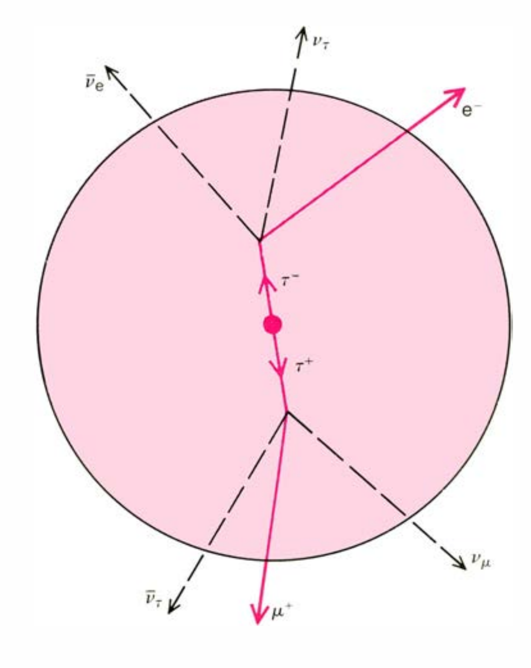
Sin embargo, en estos primeros tiempos debíamos mantener una actitud escéptica, porque con sólo 24 sucesos había varias posibilidades de que pudiéramos habernos equivocado. En primer lugar, nuestro detector distaba mucho de ser perfecto en la identificación de electrones y muones. En realidad, alrededor del 20 por ciento de las veces confundía un hadrón con un electrón o un muón. Un cuidadoso estudio del problema indicó, no obstante, que sólo cinco o seis de los 24 sucesos podían atribuirse a haber identificado erróneamente un hadrón entre los productos de desintegración.

Una segunda razón que corroboraba nuestro escepticismo era el hecho de que aun cuando hubiéramos encontrado una nueva partícula, podía tratarse de un nuevo hadrón en lugar de un nuevo leptón. Sucedía incluso que había entonces un candidato adecuado para este nuevo hadrón: el hadrón con encanto llamado mesón *D*, que aún no había sido descubierto. Era imaginable que los sucesos electrón-muón provinieran de la producción y desintegración de los mesones *D*, acompañados por neutrinos y mesones *K* neutros. El mesón *K* neutro tiene la peculiaridad de que la mitad de las veces decae tan lentamente que habría escapado más allá de nuestro detector antes de que se hubiera desintegrado. En este caso, si ambos mesones *K* neutros escaparan, se vería únicamente el electrón y el muón en el detector y se podría pensar que se ha observado realmente la desintegración de un par tau-antitau. A medida que fuimos acumulando más sucesos electrón-muón, particularmente a energías más elevadas, nuestro colega Gary J. Feldman pudo demostrar que muchos de los sucesos electrón-muón no podían provenir de la producción de pares de mesones *D*, porque escaseaban casos en los que se observara simultáneamente una señal electrón-muón y los productos esperados de la desintegración de los mesones *K* neutros.

En los dos años que siguieron al descubrimiento inicial fuimos recogiendo más sucesos electrón-muón (tenemos ahora alrededor de 200) y hemos venido considerando otros métodos para confirmar nuestros datos. Una comprobación importante era ver si la distribución en energía del electrón y el muón era la que habría de esperar de la desintegración débil de un leptón pesado en tres partículas. Encontramos que los datos se ajustaban bien a la hipótesis de una desintegración a tres cuerpos, pero que no se ajustaban la hipótesis alternativa de que se produjeran sólo dos *(véase la ilustración de la página opuesta).*

En otra comprobación consideramos otros modos posibles de desintegración del par tau-antitau. Por ejemplo, si una de las partículas tau se desintegrara en un mesón (los mesones son una subclase de los hadrones) y la otra se desintegrara en un muón, entonces sería de esperar que se vieran sucesos muón-mesón característicos. Estos sucesos serían característicos porque, igual que los sucesos electrón-muón, parecerían violar la conservación del número leptónico. Un grupo de físicos de la Universidad de Maryland, la Universidad de Pavia y la Universidad de Princeton hallaron unos cuantos de estos sucesos muón-mesón en una experiencia realizada en el SPEAR; con posterioridad, nuestro grupo Stanford-Berkeley ha sido capaz de reunir más de 100 sucesos de este tipo.



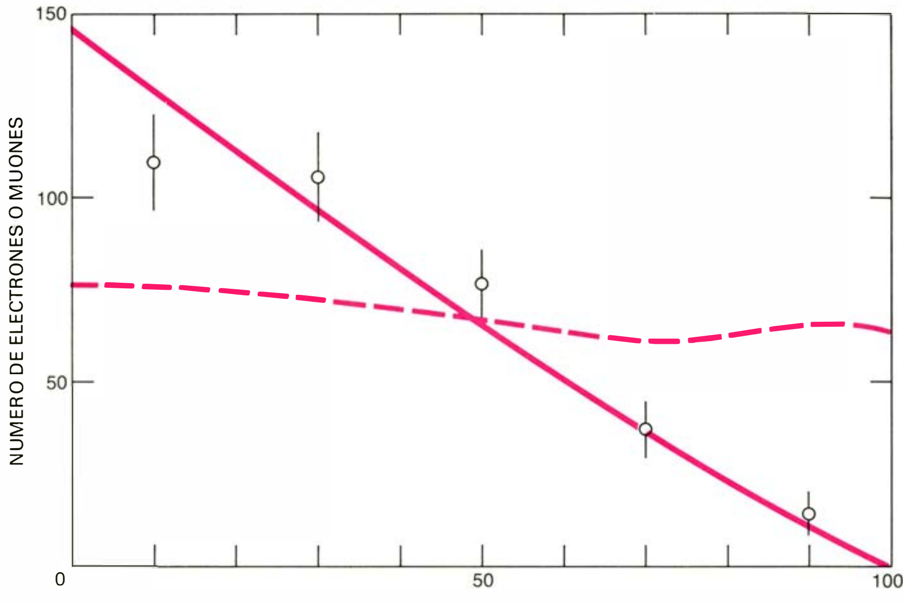


SUCESO ELECTRON-MUON TIPICO, observado recientemente en el detector de partículas cargadas del SPEAR *(izquierda).* **La señal característica de uno de estos sucesos es la detección de una partícula que atraviesa las placas de hierro (el muón) y otra partícula que no lo hace (el electrón). El suceso se interpreta en la figura de la derecha, con una gran ampliación, mediante la hipótesis de que el electrón y el muón provienen de la desintegración de un par tau-antitau. La creación y la desintegración de los leptones pesados tienen lugar en unos milímetros alrededor del centro del detector y es imposible verlos directamente. Cada partícula tau se desintegra en una partícula cargada y un par de neutrinos; sólo pueden detectarse las partículas cargadas, en este caso un electrón cargado negativamente y un muón cargado positivamente.**

Hay otros eslabones que deben cerrarse en la cadena de pruebas. El laboratorio alemán Electron Synchrotron, de Hamburgo, tiene un anillo de almacenamiento electrón-positrón llamado DORIS, dotado de una capacidad de producir partículas análogas a la del SPEAR. Si las partículas tau podían ser creadas con el aparato de Stanford, también deberían poder crearse con el de Hamburgo. Durante un año, más o menos, no hubo noticias de que se hubieran hallado leptones pesados con el DORIS; pero luego un grupo de la máquina de Hamburgo empezó a observar sucesos electrón-muón y, más tarde, sucesos muón-mesón; en ambos casos tenían las características que se esperaban para un par tau-antitau. Llegó una nueva confirmación cuando otro grupo del DORIS descubrió cierto número de sucesos electrón-mesón, que son análogos a los sucesos muón-mesón.

En una conferencia internacional sobre la física de los leptones y fotones que tuvo lugar en Hamburgo el mes de agosto del año pasado, cinco grupos experimentales que trabajaban con el SPEAR y dos grupos que lo hacían con el DORIS presentaron los resultados de sus búsquedas independientes de producción de leptones pesados en la aniquilación electrón-positrón. Todos los grupos estaban de acuerdo con los siguientes puntos: (1) que había pruebas sobre la existencia de una nueva partícula; (2) que la nueva partícula no era un hadrón con encanto; (3) que la masa de la nueva partícula estaba entre 1800 y 1900 MeV (cercana a 20 veces la masa del muón y 4000 veces la del electrón), y (4) que todas las propiedades medidas de la nueva partícula eran coherentes con las que podrían esperarse para un leptón pesado cargado.

Una forma de demostrar que la nueva partícula es un leptón y no un hadrón es medir su ritmo de producción como una función de la energía total de la colisión electrón-positrón. Estudios de la producción de pares de hadrones han demostrado que cuando la energía total del sistema aumenta por encima de una cierta energía umbral, el ritmo de producción al principio crece durante un cierto tiempo y después empieza a decrecer rápidamente. Este rápido descenso es consecuencia del hecho de que hay demasiada energía disponible y los hadrones, que son partículas compuestas, no pueden mantenerse unidas. En lugar de un par de hadrones se producen muchos hadrones. Por el contrario, como los leptones son partículas puntuales que no se pueden romper, se espera que el ritmo de producción de leptones aumente rápidamente a partir de la energía umbral, alcance un máximo y entonces empiece a decrecer muy lentamente a medida que la energía aumenta. Usando los sucesos electrón-muón como una medida del ritmo de producción de los pares tau-antitau, hemos determinado que el ritmo de producción de las nuevas partículas cambia al crecer la energía de la forma predicha para pares de leptones y no de la forma predicha para pares de hadrones *(véase la ilustración de la página siguiente)*.



PRUEBA que apoya la conclusión de que la nueva partícula elemental que se ha creado es realmente un leptón y no una nueva clase de hadrón. La ilustración representa la distribución energética de las partículas que se producen en los típicos sucesos electrón-muón. Sólo se han considerado sucesos en los que tanto el momento del muón como el del electrón eran superiores a un cierto valor (650 MeV). Suponiendo que la partícula tau es un leptón pesado que decae débilmente en tres partículas (ya sea un electrón y dos neutrinos o un muón y dos neutrinos), se espera que los datos se ajusten a la línea continua coloreada. Si la tau fuera otra clase de partícula que se desintegrara en un electrón o un muón, acompañada en ambos casos por una única partícula neutra que no se ha detectado, los datos deberían seguir la línea de trazos coloreada. La distribución energética observada es coherente con la hipótesis de tres-cuerpos, pero no con la hipótesis alternativa de dos cuerpos.

Aun cuando todos los experimentos hasta el presente están de acuerdo en que hay un nuevo leptón pesado, quedan muchas propiedades del mismo por conocer. Por ejemplo, los resultados experimentales existentes son coherentes con la idea de que la tau no queda afectada por las fuerzas fuertes, pero esta insensibilidad no ha sido comprobada tan profusamente para la tau como lo ha sido para el electrón y el muón. Además, muchas de las partículas elementales, incluidos los leptones, están en perpetua rotación alrededor de sus propios ejes, igual que peonzas, pero no ha sido establecido todavía si la tau tiene las mismas características de rotación que el electrón y el muón. Tampoco sabemos aún si la tau es el leptón secuencial pesado que buscábamos. En otras palabras, ¿tiene la tau su propio número leptónico o carácter tauónico, o es una clase, totalmente nueva, de leptones con propiedades aun más peculiares?

Experimentos realizados en otros laboratorios, basados en la interacción de neutrinos muónicos, no parecen haber creado ninguna partícula tau; por tanto, resulta poco probable que la tau tenga carácter muónico. Sin embargo, queda aún la posibilidad de que tuviera carácter electrónico. (De ser así, debe existir, a pesar de todo, un mecanismo muy especial y complicado que suprima la desintegración electromagnética de la tau en un electrón y un fotón.) Es más: ni siquiera se han observado todos los distintos modos de desintegración de la tau, y sólo conociendo el conjunto completo de los modos de desintegración se puede saber exactamente qué clase de leptón es la partícula tau.

Otro punto que exigirá nuevos estudios teóricos y experimentales es la cuestión de por qué la masa de la tau (entre 1800 y 1900 MeV) se aproxima tanto a la masa estimada del mesón D encan-

tado (1865 MeV), sobre todo ahora que dispongamos de pruebas notables de que la tau no es ningún hadrón encantado. Sucede, además, que, en física de partículas, hay otra correspondencia peculiar de esas: la masa del muón (106 MeV) es muy cercana a la masa del mesón pi o pión (140 MeV). ¿Son esto puras coincidencias o hay alguna relación desconocida entre las masas de los leptones y de los hadrones?

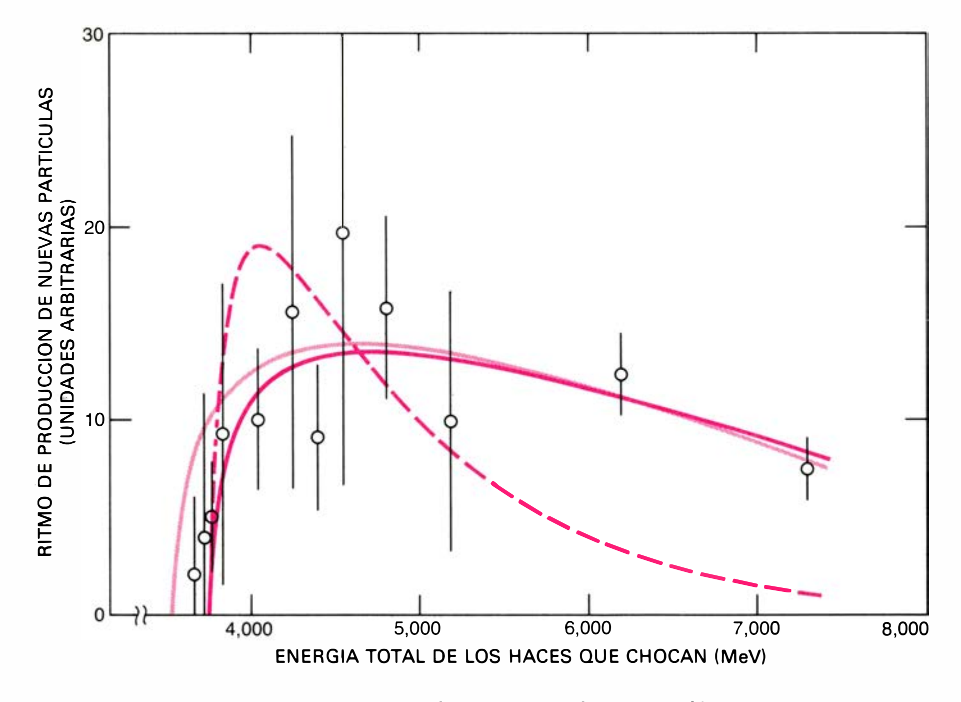
Quedan más interrogantes por despejar. Por ejemplo, ¿qué es lo que determina la masa de los leptones? Es difícil sacarle un sentido a la secuencia observada de las masas del electrón, muón y tau: 0,51, 106 y 1800 a 1900 MeV. Los números aumentan con demasiada rapidez como para ser una serie aritmética y con suma lentitud como para ser una serie geométrica. Bien es verdad que tenemos sólo tres puntos, y podemos encontrar muchas fórmulas empíricas para ajustar los números de esta secuencia, pero ninguna de ellas estará basada sobre un conocimiento fundamental de los leptones, ya que nadie sabe a qué es debida la masa de los leptones.

¿Qué debemos pensar de los caracteres electrónicos, muónico y tauónico? Quizá no haya ningún otro secreto detrás de estas propiedades. De la misma forma que los físicos han llegado a aceptar la carga eléctrica como una propiedad fundamental de las partículas, de la que no se tiene un entendimiento más profundo, quizá tendrán que aceptar también, sin más, los caracteres electrónico, muónico y tauónico de los leptones. Aún no sabemos por qué la carga eléctrica total se conserva en todas las interacciones posibles entre partículas; tal vez no es posible saber por qué los números leptónicos totales se conservan en todas las reacciones con leptones.

Sin embargo, las contestaciones pueden estar a la vuelta de la esquina. Se están construyendo en estos momentos dos nuevos aceleradores de alta energía en los que chocarán haces de electrones y positrones: la máquina PEP en el SLAC y la PETRA en Hamburgo. Ambas máquinas permitirán alcanzar una energía de 18.000 MeV o más en cada haz; ello significa que la búsqueda de nuevos leptones cargados se podrá extender hasta masas quizá cinco veces mayores que el límite superior actual, fijado en unos 3500 MeV. Muchas de las experiencias que están siendo planeadas para el PEP y el PETRA incluyen en su propuesta formas para buscar nuevos leptones pesados, normalmente mediante la observación de sucesos electrón-muón. Tanto si se encuentran o no nuevos leptones, su búsqueda será más difícil, aunque sólo sea porque los sucesos electrón-muón que provienen de la desintegración de la tau constituirán un fondo molesto. Los experimentales tendrán que distinguir los nuevos e interesantes sucesos electrón-muón de aquellos otros viejos y poco interesantes que provendrán de las desintegraciones del par tau-antitau. Por otra parte, los sucesos electrón-muón originados por las tau serán útiles en la puesta a punto de las nuevas máquinas.

Finalmente, hay una cuestión más profunda para la que no disponemos todavía de ninguna respuesta. ¿Cuál es la relación entre las dos clases de partículas con varios miembros que ahora parecen ser realmente fundamentales, los leptones y los quarks? Antes del descubrimiento de la partícula tau había únicamente cuatro tipos conocidos de leptones y cuatro tipos conocidos de quarks (contando cada partícula y su antipartícula como un solo tipo). Había, por tanto, una admirable simetría, y ciertas teorías que relacionaban los leptones con los quarks hacían uso de ella. Con el descubrimiento del nuevo leptón pesado la simetría se ha roto; hay ahora más leptones conocidos que quarks conocidos, dos más si la tau es un leptón secuencial con su propio neutrino tauónico.

Sin embargo, un grupo de físicos del Fermi National Accelerator Laboratory (Fermilab) ha presentado recientemente pruebas de la posible existencia de una nueva y quinta clase de quarks. Así pues, el número de quarks puede también estar creciendo. Algunas teorías preservan la simetría entre los leptones y los quarks, mientras que otras renuncian a esta simetría, pero casi todas admiten la posibilidad de nuevos quarks y leptones. Esta aparente proliferación en los tipos de leptones y quarks, aunque sea aún una materia de especulación, es de alguna forma alarmante. En muchos sentidos, sería preferible que las partículas realmente elementales pudieran ser tan cortas en número como lo eran en los días anteriores al descubrimiento de la tau, o quizá mejor a los tiempos anteriores a la tau y al quark encantado. Obviamente, no se puede dictar a la materia cuáles deberían ser sus constituyentes fundamentales. Sólo podemos esperar que seamos capaces de encontrar tales constituyentes y entender sus propiedades.



NUEVA PRUEBA de que la tau es un leptón y no un hadrón. En la gráfica se representa el ritmo de producción de la nueva partícula como función de la energía total liberada en el choque electrón-positrón. Si las nuevas partículas fueran hadrones, entonces, a medida que aumenta la energía total del sistema por encima de una determinada energía umbral, su ritmo de producción se esperaría que creciera durante un cierto tiempo y que después decreciera rápidamente *(línea de puntos coloreada).* **Si las partículas tau son leptones su ritmo de producción debería crecer rápidamente a partir de la energía umbral, alcanzar un máximo y después decrecer lentamente a medida que aumenta la energía** *(líneas coloreadas continuas).* **Los datos son coherentes con la hipótesis de un leptón-pesado y no con la hipótesis alternativa. La línea coloreada obscura se basa en la hipótesis de que la masa de los nuevos leptones es 1900 MeV; la clara, en la hipótesis de que la masa del nuevo leptón es 1800 MeV. Los datos experimentales sugieren que la masa real está entre estos dos valores.**

La superficie de Marte

Los vehículos Viking han suministrado de ella una perspectiva inigualada desde la órbita y desde el suelo, incorporando mucha información sobre el modo como los volcanes, impactos meteoríticos, agua y viento le han dado forma

Raymond E. Arvidson, Alan B. Binder y Kenneth L. Jones

Los dos módulos de descenso Viking llevan ya en la superficie de Marte casi un año marciano completo; los dos módulos orbitales han estado dando vueltas y fotografiando el planeta durante algo más. Desde que se posaron los módulos de descenso, en el verano de 1976, han estado recogiendo datos sobre las características de la atmósfera marciana, rocas y suelo. Mientras tanto, los vehículos orbitales han estado vigilando la concentración de vapor de agua en la atmósfera, la distribución de temperatura en la superficie y han fotografiado ésta con claridad y resolución desconocidas hasta ahora.

El conjunto de las fotografías tomadas y de los experimentos analíticos realizados por los cuatro vehículos espaciales revela que Marte es un planeta con una historia todavía más complicada de lo que se sospechaba. Tal conjunto ha suministrado pruebas de que incluso el más antiguo terreno picado de cráteres ha sido modificado por la actividad volcánica, de que en la temprana historia del planeta las corrientes de agua constituían un factor importante para conformar sus características y de que, desde entonces, la materia de su superficie ha sido redistribuida considerablemente por vientos de gran velocidad. Sorprendentemente, la superficie ha sido poco erosionada por tal actividad eólica. El aspecto de la superficie de Marte se aproxima más a los desiertos de rocas volcánicas de la Tierra que a la superficie densamente salpicada de cráteres de la Luna; pese a ello, Marte, que antiguamente se imaginaba ser en gran parte un mundo de dunas suavemente onduladas, parece poseer poca arena. La misión Viking también ha suministrado datos que en parte confirman, y en parte alteran, hipótesis acerca de la primitiva atmósfera y clima de Marte formuladas tras la misión *Mariner 9* de 1971 y 1972.

El *Mariner 9* fotografió prácticamente la totalidad de la superficie de Marte con una resolución de unos kilómetros, y una pequeña porción de la superficie con una resolución de cientos de metros. La cobertura global de Marte por el *Mariner 9* indicó que el planeta estaba dividido en dos hemisferios claramente diferentes: un hemisferio sur escarpado, fuertemente salpicado de cráteres y atravesado por grandes depresiones acanaladas, y un hemisferio norte liso, con sólo pocos cráteres y salpicado de volcanes extinguidos. Las dos clases de terreno están aproximadamente separadas por un círculo máximo inclinado unos 30 grados respecto al ecuador.

Marte antes de los Viking

Los cráteres del hemisferio sur tienen tamaño variable hasta alcanzar el de la depresión Hellas, de 1600 kilómetros de diámetro. La abundancia de cráteres en algunas de las regiones más densamente salpicadas es comparable a la que se encuentra en las altiplanicies claras de la Luna. Se han determinado las edades de muestras de rocas y terreno recogidas en las altiplanicies lunares durante las misiones Apolo, y tales edades indican que la mayoría de los cráteres de las altiplanicies lunares se originaron entre hace 4000 y 4500 millones de años. En tal época, la Luna fue bombardeada torrencialmente por los restos de materia interplanetaria no utilizados en la formación del sistema solar. Como la abundancia de cráteres en Marte es similar a la de la Luna, se cree que el terreno marciano cubierto de cráteres es aproximadamente de la misma edad que las altiplanicies lunares. Dicho de otra forma, casi la mitad de la superficie de Marte está constituida por terreno antiguo en el que muchos de los accidentes importantes del mismo se han mantenido inalterados durante unos 4000 millones de años.

El *Mariner 9* reveló que las áreas escasamente picadas del hemisferio norte de Marte son llanuras de lava que inundó ampliamente la superficie después de haber cesado el intenso bombardeo del planeta. Los pocos cráteres que hay en las llanuras registran el impacto ocasional de un asteroide o cometa suelto. Aunque todavía no hay manera fidedigna de estimar las edades absolutas de las llanuras, la abundancia de cráteres en diferentes regiones varía ampliamente, lo que implica que las edades de las llanuras oscilan desde cientos a miles de millones de años.

Los canales de Marte, tan detalladamente fotografiados por el *Mariner 9*, parecían indicar que el clima del planeta en el pasado había sido muy diferente del actual. Si toda el agua de la atmósfera marciana existente hoy se condensara en un solo lugar, produciría una masa de agua no mayor que el estanque Walden. De hecho, la abundancia de agua en la atmósfera actual es tan pequeña que la lluvia resulta absolutamente imposible.

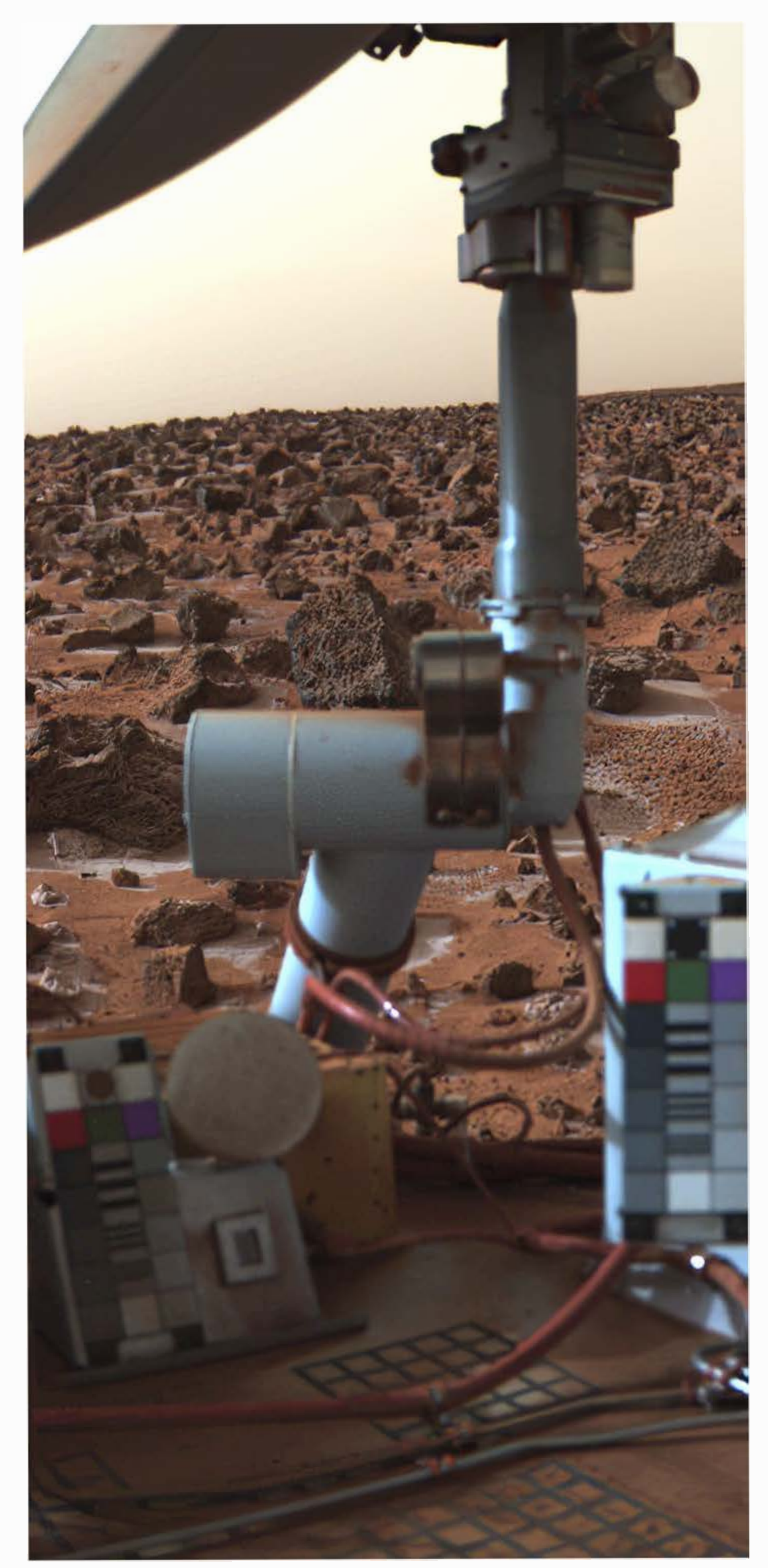
Por la época en que se posaron los módulos de descenso Viking, la mayoría de los investigadores estaban convencidos de que los canales mayores, cuya anchura es de decenas de kilómetros, habían sido excavados por agua procedente de la fusión de hielo debajo de la superficie. Algunos investigadores consideran que el hielo es un residuo de una atmósfera más densa que el planeta pudo tener durante su primer millar de millones de años. Según cálculos realizados por Fraser P. Fanale, del Laboratorio de Propulsión a Chorro del Instituto de Tecnología de California, la atmósfera marciana primitiva pudo contener amoníaco y metano, además de dióxido de carbono y vapor de agua. Según James B. Pollack, del Centro de Investigación Ames de la Administración Nacional de Aeronáutica y del Espacio (NASA), tal

atmósfera habría retenido la radiación infrarroja ("efecto de invernadero") y habría estado suficientemente caliente para contener en su seno cantidades importantes de vapor de agua. En algún momento del pasado, diversas reacciones químicas habrían eliminado el amoníaco y el metano de la atmósfera. Esta se habría vuelto entonces más transparente a la radiación infrarroja. Como consecuencia, la atmósfera se habría enfriado y el vapor de agua contenido en ella se habría condensado y precipitado sobre la superficie, donde se habría dirigido hacia las regiones polares y hacia el interior de la corteza del planeta, fracturada por los impactos, introduciéndose también en la regolita (materia suelta que descansa sobre la corteza).

El *Mariner 9* reveló también que tanto el polo norte como el sur de Marte están cubiertos de hielo y de polvo arrastrado por el viento en un espesor de varios kilómetros. En los depósitos más antiguos, el hielo y el polvo no están distribuidos en capas; en los más modernos, se alternan en capas, cada una de varias decenas de metros de espesor.

Las grandes cantidades de polvo existentes en esos depósitos suministran otra prueba de que, en cierta época, la atmósfera de Marte fue más densa de lo que es hoy; la atmósfera actual no parece capaz de transportar tales cantidades de materia a los polos. De hecho, el proceso más reciente ha consistido en una erosión parcial de los depósitos por el viento, trasladando desde ellos materia hasta formar una capa que recubre gran parte de la superficie en latitudes altas. Los depósitos estratificados implican que el antiguo clima de Marte no solamente fue diferente, sino que pudo haber estado sometido a cambios periódicos.

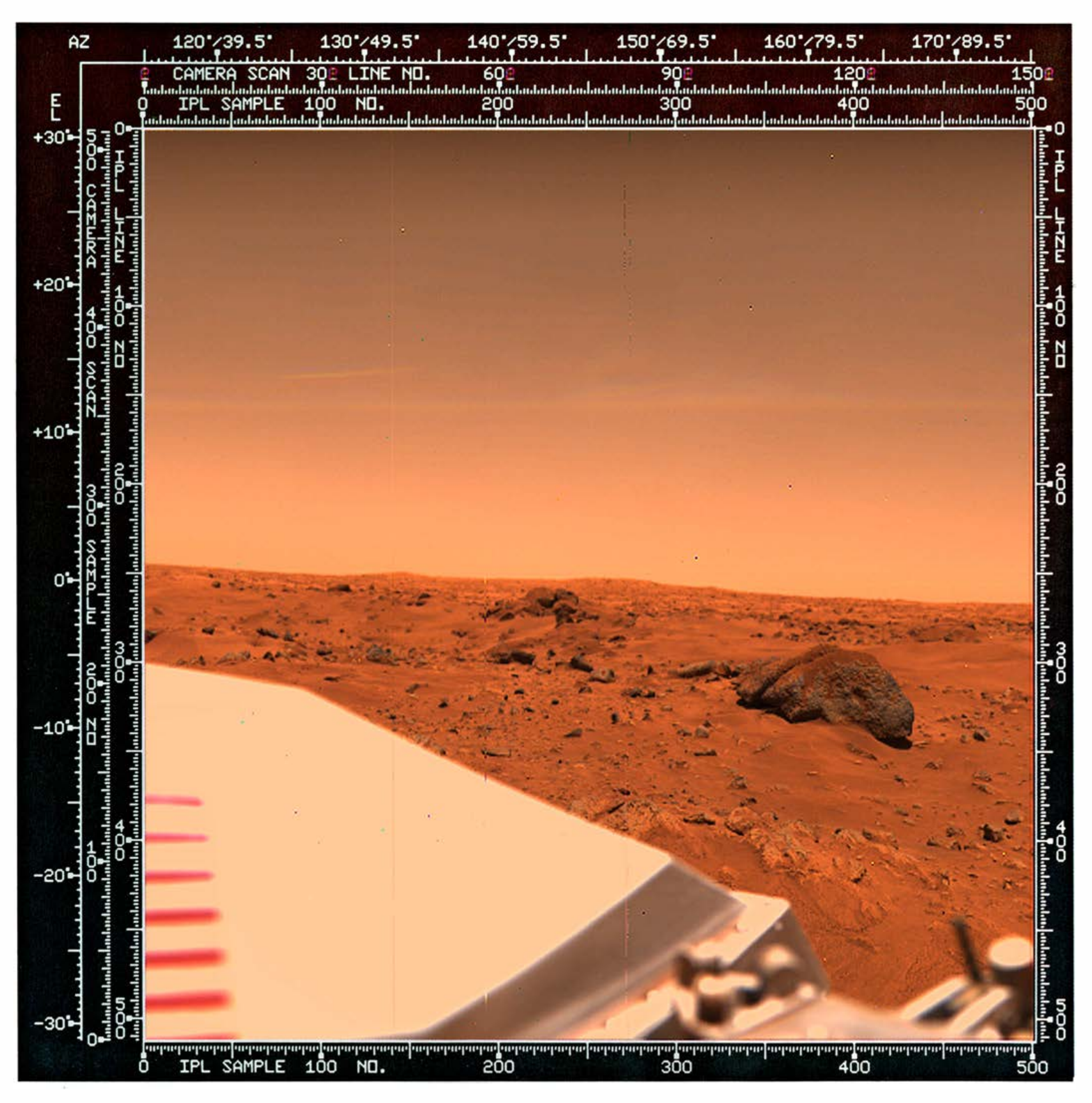
Basándose en las fotografías del *Mariner 9,* se propusieron varias hipótesis sobre el origen de los depósitos polares.



ESCARCHA EN MARTE, fotografiada por el módulo de descenso *Viking 2* cerca del mediodía local del 13 de septiembre del año pasado, lo que corresponde a finales de invierno en el hemisferio norte marciano. El módulo se posó en Utopia Planitia a una latitud de 48 grados. La vista se tomó mirando al suroeste por encima de la parte superior del vehículo. La escarcha son las manchas blancas en la base de las rocas. Probablemente son los restos de un fino depósito de verdadero hielo, quizá mezclado con un clathrato en que la nieve carbónica queda encerrada en el interior del hielo. La escarcha se había acumulado lentamente a principios de invierno, a medida que se transportaba agua al hemisferio norte desde el hemisferio sur. El análisis del color de la superficie indica que la escarcha había estado sobre el suelo durante al menos 100 días antes de tomar esta fotografía.

Algunos investigadores sugirieron que los depósitos de hielo y polvo se acumularon principalmente durante el período primitivo, cuando la atmósfera de Marte era más densa. Otros manifestaron que los depósitos polares se habían estado acumulando durante la totalidad de la historia marciana y que su acumulación ha sido modulada por la cantidad de radiación solar recibida por el planeta, particularmente en los polos. Los cálculos de Carl Sagan y sus colegas, de la Universidad Cornell, y los de William K. Hartmann, del Instituto de Ciencias Planetarias de Flagstaff, Arizona, indicaron que la cantidad de calor recibida por Marte podría haber variado en la cantidad requerida si la luminosidad del Sol se hubiera alterado. Se cree que tales alteraciones se producen a lo largo de un período que va de un millón a cien millones de años.

Además, William R. Ward, del Laboratorio de Propulsión a Chorro, demos-



"PEPE EL GORDO" es el nombre con que se bautizó la roca grande y obscura a la derecha del centro de esta fotografía tomada por el módulo de descenso *Viking 1* a mediodía local el 23 de agosto de 1976. La roca tiene unos dos metros de diámetro y está a unos 10 metros del módulo; es una de las varias rocas que se pueden ver reposando entre montoncillos de material de grano fino. Es probable que Pepe el Gordo estuviera en alguna época cubierto de material arrastrado por el viento, parte del cual todavía parece adherido a la parte superior en forma de depósito pardo amarillento. La película se tomó en el verano del hemisferio norte, estación caracterizada por su falta de tempestades de polvo, pese a lo cual el color pardo amarillento del cielo se debe a partículas de polvo suspendidas. Parece que en la atmósfera de Marte hay siempre algo de polvo en suspensión. En el centro de la fotografía puede verse el perfil de un cráter de impacto a unos 300 metros del módulo. Las cifras en el margen de la fotografía ayudan en la interpretación de ésta. Las cifras de la columna vertical más a la izquierda indican la elevación en grados por encima y debajo del plano horizontal de la cámara; las cifras en las dos filas horizontales de más arriba indican el azimut en grados, en el sistema coordenado de la cámara y en el del módulo, respectivamente. El norte exacto está a un azimut de 130 grados. Los otros grupos de cifras que aparecen también en la fotografía se refieren a las líneas de exploración mediante las cuales la cámara construyó la imagen.

tró que la inclinación del eje de rotación de Marte cambia considerablemente con el tiempo a causa de la mayor atracción del Sol sobre el ensanchamiento ecuatorial del planeta. En la actualidad, el eje de rotación de Marte está inclinado unos 25 grados respecto a la perpendicular al plano de su órbita. A lo largo de un período de entre 100.000 y un millón de años, sin embargo, la inclinación del eje cambia desde un mínimo de 15 grados hasta un máximo de 35. Para complicar las cosas todavía más, cálculos más recientes debidos a Ward, Joseph A. Burns, de la Universidad de Cornell, y O. Brian Toon, del Centro de Investigación Ames, muestran que la inclinación del eje podría ocasionalmente haber alcanzado 45 grados durante un período primitivo, antes de la formación de la gran prominencia volcánica Tharsis.

Un radiómetro instalado a bordo del primitivo vehículo espacial *Mariner 7* reveló que los casquetes de hielo estacionales de Marte estaban constituidos por anhídrido carbónico congelado (nieve carbónica). Parecía probable que el casquete de hielo residual que quedaba durante el verano estuviera también compuesto de nieve carbónica. Si ésta constituía el casquete permanente, un aumento relativamente pequeño de la temperatura atmosférica debería incrementar la presión de la atmósfera marciana desde su valor actual, de 2 a 10 milibares, hasta otro que se hallara entre 30 milibares y un bar, que es la presión atmosférica de la atmósfera terrestre al nivel del mar. La estima de 30 milibares, calculada por Bruce C. Murray y sus colegas, del Instituto de Tecnología de California, se basaba en la cantidad de radiación solar recibida por los casquetes de hielo cuando el eje de rotación de Marte estuviera a su inclinación máxima de 35 grados, y en la hipótesis de que la atmósfera transportaba una cantidad mínima de calor a los polos. La estima de un bar, calculada por Sagan y sus colaboradores, era un límite superior resultante de un efecto de invernadero realimentado, mediante el cual un aumento inicial de la temperatura conduce a un aumento de la presión atmosférica, que permite un mayor transporte de calor a los polos, lo que a su vez permite que se evapore más nieve carbónica.

En períodos más cálidos, o durante aquellos en los que los polos de Marte apuntaran hacia el Sol más de lo que lo hacen en la actualidad, la densidad de la atmósfera marciana habría sido mayor, lo que habría erosionado una mayor cantidad de materia polvorienta en las bajas latitudes, que habría sido transportada hacia los polos. A medida que la atmósfera se enfriaba en los polos y se condensaba para formar el casquete de hielo, el polvo suspendido en ella se habría depositado junto con la materia condensada. En los períodos más fríos, sin embargo, se habría condensado en los polos una porción tan grande de la atmósfera, que a latitudes más bajas la porción gaseosa restante estaría muy enrarecida y transportaría mucho menos polvo hacia los polos.

Las observaciones de los módulos orbitales Viking

Los módulos orbitales Viking, que comenzaron su cobertura de Marte a principios del verano de 1976, fotografiaron el planeta con mucho mayor claridad que el *Mariner 9*. La diferencia se debía en gran parte a que el *Mariner 9* se aproximó a Marte cuando el planeta estaba todavía envuelto en una tempestad de polvo. Además, pasada la tempestad, quedó suspendido en la atmósfera polvo suficiente para enturbiar considerablemente la vista de la superficie de Marte durante varios meses terrestres. Los módulos orbitales Viking comenzaron a tomar imágenes de Marte cuando la atmósfera estaba relativamente limpia de polvo. Y aunque las cámaras de los módulos orbitales Viking eran sistemas Vidicon análogos a los que llevaba el *Mariner 9*, poseían una resolución apreciablemente mejor.

Las fotografías de los módulos orbitales Viking muestran que gran parte de la superficie de Marte conserva marcados rasgos topográficos; se distinguen claramente ríos de lava, crestas de plegamiento y materia eyectada por los cráteres. Además, las fotografías muestran numerosos ríos de lava incluso en el suelo con cráteres más primitivo. Basándose en esos últimos rasgos, Michael H. Carr, del Servicio Geológico de los Estados Unidos, y sus colegas sugieren que, al principio de la historia de Marte, la lava llegó a inundar gran parte del terreno antiguo salpicado de cráteres, en el hemisferio sur. Tal nivelado volcánico explicaría por qué el terreno de la zona de Marte abundantemente salpicada de cráteres es relativamente liso en comparación con las montañosas altiplanicies de la Luna.

El que los más antiguos accidentes de Marte estén todavía agudamente definidos indica también que, a lo largo de la historia del planeta, ha habido relativamente poca fractura de rocas y redistribución de residuos. La única prueba clara de erosión en gran escala producida por el viento se encuentra en las regiones constituidas por los depósitos sedimentarios más antiguos, como son las regiones cercanas a los polos. Es probable que los depósitos polares estén sólo parcialmente aglomerados y que, por tanto, resulten fácilmente erosionados por el viento.

Los módulos orbitales Viking han puesto de manifiesto que las llanuras septentrionales de las altas latitudes son algo más que simples aglomerados de residuos. Forman una compleja mezcla de corrientes de lava y depósitos de polvo arrastrado por el viento, que a su vez han sido también limados por el viento. Más cerca del polo norte, las cámaras de los vehículos orbitales hallaron grandes campos de dunas ciñendo el casquete de hielo residual; las dunas parecen estar constituidas por partículas del tamaño de granos de arena. Estos granos, que en Marte probablemente alcanzan desde una décima de milímetro a varios milímetros de diámetro, son demasiado grandes y pesados como para permanecer suspendidos en una atmósfera tan enrarecida y para ser llevados con el viento; sin embargo, son lo suficientemente pequeños y ligeros como para que éste los haga rodar y los levante cortas distancias. Pero las partículas de polvo, que en Marte no llegan a una décima de milímetro sí son suficientemente pequeñas y ligeras para ser arrastradas en suspensión.

Tanto en el caso de los residuos de las llanuras septentrionales como en el de las dunas más próximas a los polos, la porción substancial parece estar formada por material erosionado de los propios depósitos polares. Cuando los depósitos fueron erosionados por el viento, el polvo incorporado en ellos fue arrastrado, mientras que la arena quedó atrás y se acumuló en dunas cercanas a los polos. La dificultad que encontramos aquí es que, aunque los depósitos polares contengan polvo, parecen estar suministrando polvo y arena. La solución puede estribar en que las partículas de tamaño de granos de arena sean en realidad partículas del tamaño del polvo que hayan quedado soldadas entre sí por óxidos, sales o quizás incluso hielo.

La interpretación de los datos de los radiómetros montados en los vehículos orbitales, realizada por Hugh H. Kieffer, de la Universidad de California en Los Angeles, y sus colaboradores, mostró que durante el verano del hemisferio norte de Marte, la temperatura sobre el casquete polar residual era de unos 205 grados Kelvin (–68 grados Celsius). Este resultado era sorprendente. Como la presión atmosférica en la superficie marcia-

na es sólo de unos seis milibares, la temperatura tendría que haber sido inferior a 148 grados Kelvin para mantener un casquete permanente de nieve carbónica. Aun cuando el hielo en el casquete polar fuese de textura clathrática, en la que el dióxido de carbono sólido estuviera encerrado en el interior de hielo verdadero, no podría existir a temperaturas superiores a unos 155 grados Kelvin. El único producto condensado que puede permanecer estable a 205 grados Kelvin es hielo verdadero, procedente de agua.

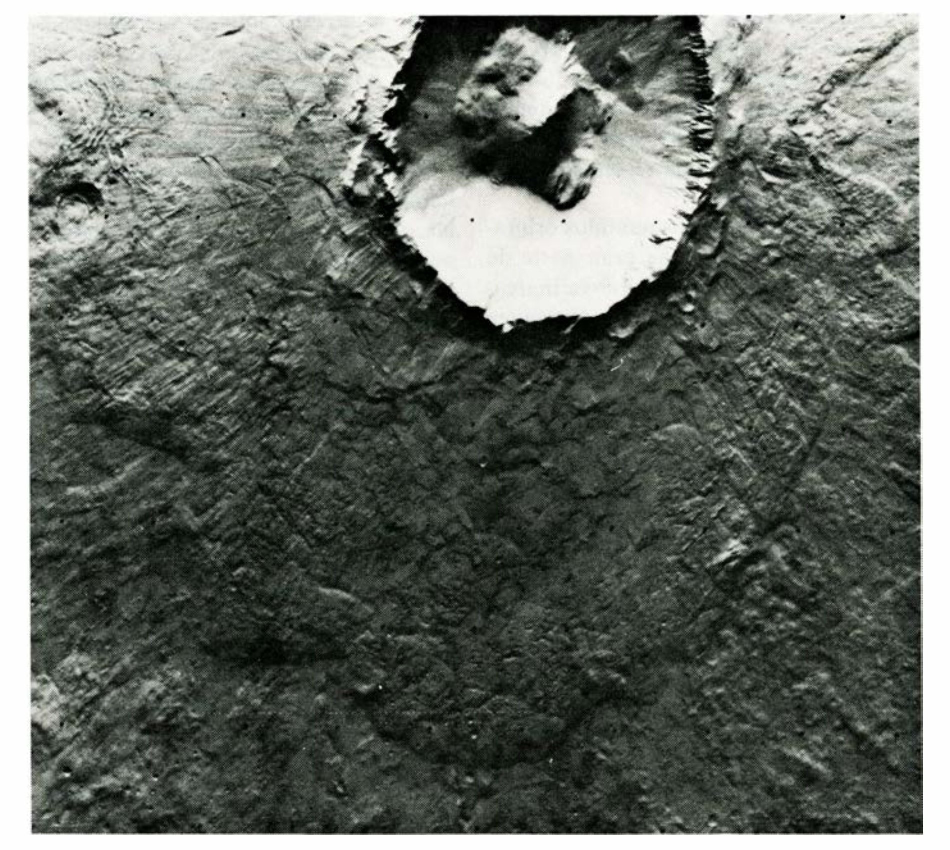
Una comprobación adicional que apoya la idea de que el casquete polar residual está formado sólo de hielo verdadero proviene de un análisis de los datos de los espectrómetros de los módulos orbitales Viking realizado por Crofton B. Farmer y sus colegas en el Laboratorio de Propulsión a Chorro. Ellos han hallado que, durante el verano, en las latitudes septentrionales la cantidad de vapor de agua en la atmósfera es tal que la temperatura debe exceder los 200 grados Kelvin.

Cuesta interpretar las observaciones del casquete polar residual en el hemisferio sur, en verano, debido a la interposición de una nube de polvo global que puede haber modulado las temperaturas atmosféricas. El casquete polar residual, meridional, consta probablemente también de hielo verdadero, si bien las medidas de su temperatura son ambiguas y existe una pequeña probabilidad de que su componente principal sea nieve carbónica.

Puesto que, según parece, hay hielo en ambos polos, las hipótesis que incorporan el agua como agente activo en el pasado de Marte han ido ganando aceptación gradualmente. Además, las fotografías de los módulos orbitales Viking han revelado que los canales de Marte son todavía más abundantes de lo que indicaban las fotografías del *Mariner 9*, extendiéndose hasta tamaños menores y formando un sistema de drenaje mucho más integrado de lo que previamente se había percibido. El escurrimiento de la lluvia procedente de una primitiva y densa atmósfera puede haber sido capaz de excavar algunas de las redes arborescentes de canales. Otros pueden haberse formado al fundirse hielo subterráneo a causa del calor desprendido por actividad volcánica. En este caso, la fusión del hielo provocaría el desplome del terreno que se hallaba encima, creando así los canales a lo largo de los que habría fluido el agua. La formación de estos canales, por tanto, estaría directamente relacionada con la historia térmica del planeta.

Una comprobación intrigante de la presencia de agua en la corteza y en la regolita de Marte la proporciona el examen de fotografías de los peculiares terraplenes, murallas y promontorios característicos de la materia eyectada en muchos de los cráteres grandes. En la Luna, la materia eyectada de los cráteres producidos por impactos parece haber sido bloques de material que fueron proyectados hacia el exterior por el impacto original y que después cayeron a la superficie, excavando a su vez miriadas de cráteres de impacto secundarios. En Marte, los depósitos de materia eyectada rodean los cráteres de impacto casi como un flujo solidificado. La explicación probable es que el calor desarrollado en el impacto fundió y vaporizó el hielo que estaba ocluido en la corteza; el agua líquida y el vapor transportaron el material eyectado hacia el exterior del cráter, constituyendo una corriente generalizada que se movía a favor de los accidentes del terreno. En algunas fotografías puede verse cómo la corriente bordeó los obstáculos que aparecían en su camino.

Aunque se va reconociendo ya que el agua ha sido un importante agente en el pasado de Marte, el descubrimiento de que las regiones polares están probablemente dominadas por hielo verdadero en lugar de nieve carbónica establece serias limitaciones a la intensidad de las pasadas fluctuaciones climáticas. Si un casquete de hielo de Marte recibiera más radiación solar, comenzaría a evaporarse, sublimándose directamente de hielo sólido a vapor. Para que la presión de vapor de agua alcanzara en Marte el valor requerido para la existencia del agua líquida, la temperatura tendría que haberse elevado al menos 70 grados Kelvin. Tan fantástico aumento de temperatura es extremadamente improbable, aun cuando la radiación solar variase en la máxima cuantía permitida por la teoría o la inclinación del eje del planeta cambiase periódicamente a su máxima amplitud. En realidad, el hecho de que las superficies de Marte más antiguas se hallen tan bien conservadas concuerda con la hipótesis de que las condiciones atmosféricas del planeta no han fluctuado grandemente a lo largo de su historia. Parece probable que la parte principal de los depósitos polares se formase muy



EL FLUJO DE RESIDUOS que rodea el cráter Arandas de 25 kilómetros, fotografiado por el módulo orbital *Viking 1* el 22 de julio de 1976. Aunque el borde del cráter está claramente definido, el material alrededor de él parece haber fluido sobre el suelo en lugar de haber sido proyectado cuando un meteorito chocó con la superficie. En la parte superior izquierda, el flujo puede haber sido desviado en torno a un cráter más pequeño. Los surcos radiales de la superficie del flujo pueden haber sido marcados en ella durante las últimas fases del proceso de impacto. El flujo se produjo quizá cuando el calor desprendido en el impacto del meteorito fundió el hielo bajo la superficie; el agua y el vapor transportaron material hacia fuera del cráter en una corriente ordenada.

pronto y que, desde entonces, hayan sido erosionados por el viento. La época exacta en que se formaron y la razón para su formación, junto con la historia de cualquier atmósfera densa primitiva que pueda haberse formado a favor de los efectos de invernadero, siguen siendo, sin embargo, misterios.

El módulo de descenso *Viking 1* se posó en las laderas occidentales de la región llamada Chryse Planitia (22,5 grados norte de latitud, 47,8 grados oeste de longitud). Desde la órbita, el lugar de toma se parece mucho a la superficie de uno de los mares lunares: se trata de una llanura volcánica lisa, salpicada de escasos cráteres y arrugada por una serie de cadenas de plegamiento. Las paredes de los cráteres parecen casi nuevas, y las crestas están asimismo bien conservadas. La cantidad de erosión debe ser muy baja o estar confinada a escalas del orden de metros para que la morfología de los depósitos eyectados y de las crestas haya cambiado tan poco.

El lugar de toma se encuentra a unos 130 kilómetros al este de Lunae Planum, una de las llanuras de Marte más densamente poblada de cráteres. Lunae Planum está aproximadamente un kilómetro más alta que Chryse Planitia, y el límite entre ambas regiones queda marcado por una escarpadura irregular. Cierto número de canales atraviesan Lunae Planum, emergen de la escarpadura y se extienden hacia el este a través de Chryse Planitia hacia el lugar de descenso. Con toda probabilidad, los canales se formaron a causa del agua procedente del hielo ocluido en el interior de Lunae Planum. En algún momento del pasado, el volcanismo y el calor interno fundieron parte del hielo subterráneo y el agua escapó hacia la superficie para crear una o varias inundaciones torrenciales que se abrieron camino a través de los depósitos de Lunae Planum y se precipitaron sobre Chryse Planitia. El flujo de las aguas de Lunae Planum abrió brechas en varios de los cráteres y crestas al oeste del punto de asentamiento.

La superficie de Chryse Planitia

A juzgar por las correlaciones entre fotografías tomadas por el módulo de descenso *Viking I* y por los módulos orbitales Viking, el módulo de descenso reposa en la falda de una cresta. Desde dicho módulo, el punto de descenso se parece sorprendentemente a muchos desiertos rocosos de la Tierra, en particular a aquellos en que hay rocas volcánicas al descubierto. El paisaje, suavemente ondulado, tiene un color castaño amarillento, salpicado de rocas y sembrado de dunas formadas por material de grano fino. A 30 metros del vehículo espacial se pueden discernir varios afloramientos del lecho rocoso. No pueden apreciarse en las fotografías del módulo de descenso indicios inequívocos de la inundación procedente de Lunae Planitia: no se detectan características suavizadas, canales o depósitos fluviales. Al parecer, la inundación o bien no alcanzó el punto de descenso o había perdido gran parte de su intensidad cuando pasó por él. Otra posibilidad es que la superficie haya sido modificada después de la inundación hasta el punto de hacer indiscernible cualquier aspecto fluvial.

La semejanza entre el punto de descenso y los desiertos rocosos de la Tierra constituyó una sorpresa, teniendo en cuenta lo que la mayor parte de los investigadores esperaban fundados en las fotografías orbitales, de aspecto lunar, de Chryse Planitia. Desde la órbita, los rasgos dominantes de la región son los cráteres. Desde el suelo, sólo se pueden ver claramente unos pocos cráteres en la inmediata vecindad del módulo. Basándose en la población de grandes cráteres visible desde la órbita, Edward A. Guinness, Jr., de la Universidad de Washington, calculó que si Marte fuese como la Luna, debería haber en el campo de visión del módulo de descenso unos 35 cráteres de diámetros comprendidos entre 25 y 50 metros.

En realidad, la existencia de un déficit de cráteres de menos de 50 metros de diámetro en Marte había sido predicha, en 1970, por Donald E. Gault y Barrett S. Baldwin, Jr., del Centro de Investigación Ames. Estos investigadores calcularon que, pese a su delgadez, la atmósfera marciana es suficientemente densa para erosionar y romper en pedazos los meteoritos pequeños incidentes, antes de que alcancen la superficie del suelo. Resulta así que Marte no está sometido al repetido impacto de pequeños objetos de gran velocidad y la consiguiente "roturación" de los pocos metros superiores de la superficie. La única población importante de cráteres de diámetros inferiores a 50 metros debe ser la de los cráteres secundarios producidos por el impacto de materiales expelidos durante la formación de un cráter grande (es decir, de decenas de kilómetros de diámetro). En contraste, la superficie de la Luna muestra un espectro continuo de tamaños de cráteres, y el impacto de pequeños objetos a lo largo de millones de años ha creado un suelo polvoriento. En Marte, los impactos grandes fracturarían la superficie y repartirían sobre ella una capa discontinua de bloques relativamente grandes, tal como se encuentra en el punto de descenso. El suelo de Marte se debe haber formado mediante otros procesos.

Además de las rocas y afloramientos visibles sobre el suelo en Chryse Planitia, se encuentra material de grano fino en abundancia en forma de regueros o vetas al lado de sotavento de la mayor parte de las rocas; los regueros tienen varios centímetros de profundidad y de 10 centímetros a un metro de largo. Al nordeste del módulo hay un conjunto de montones en un campo de grandes guijarros; los montones probablemente se formaron al ser detenido el material, arrastrado por el viento, por guijarros que eran suficientemente grandes como para reducir localmente la velocidad del viento. Varios de los grandes montones presentan apariencia laminada o estratificada. En la Tierra no suelen verse estratos en dunas o amontonamientos que están creciendo activamente o moviéndose; solamente se hacen visibles cuando el material de arrastre ha sido estabilizado por la vegetación o por cementación y está siendo erosionado por el viento. Parece que, en Marte, los amontonamientos observados se despositaron hace algún tiempo, se litificaron parcialmente (es decir, se transformaron en roca sedimentaria) y han sido recientemente erosionados.

Por término medio, los ejes mayores de las vetas arenosas situadas detrás de las rocas apuntan casi exactamente hacia el sur. Fotografías tomadas desde el *Mariner 9* en regiones un poco al norte del punto de descenso muestran grandes vetas que parten de cráteres y se extienden también en general hacia el sur. Ambos conjuntos de vetas parecen indicar que los vientos dominantes en la proximidad de la superficie habían soplado de norte a sur durante el período de formación de las vetas. Además, la configuración de los estratos visibles en las dunas indican que cuando éstos se depositaron, la dirección del viento era también de norte a sur.

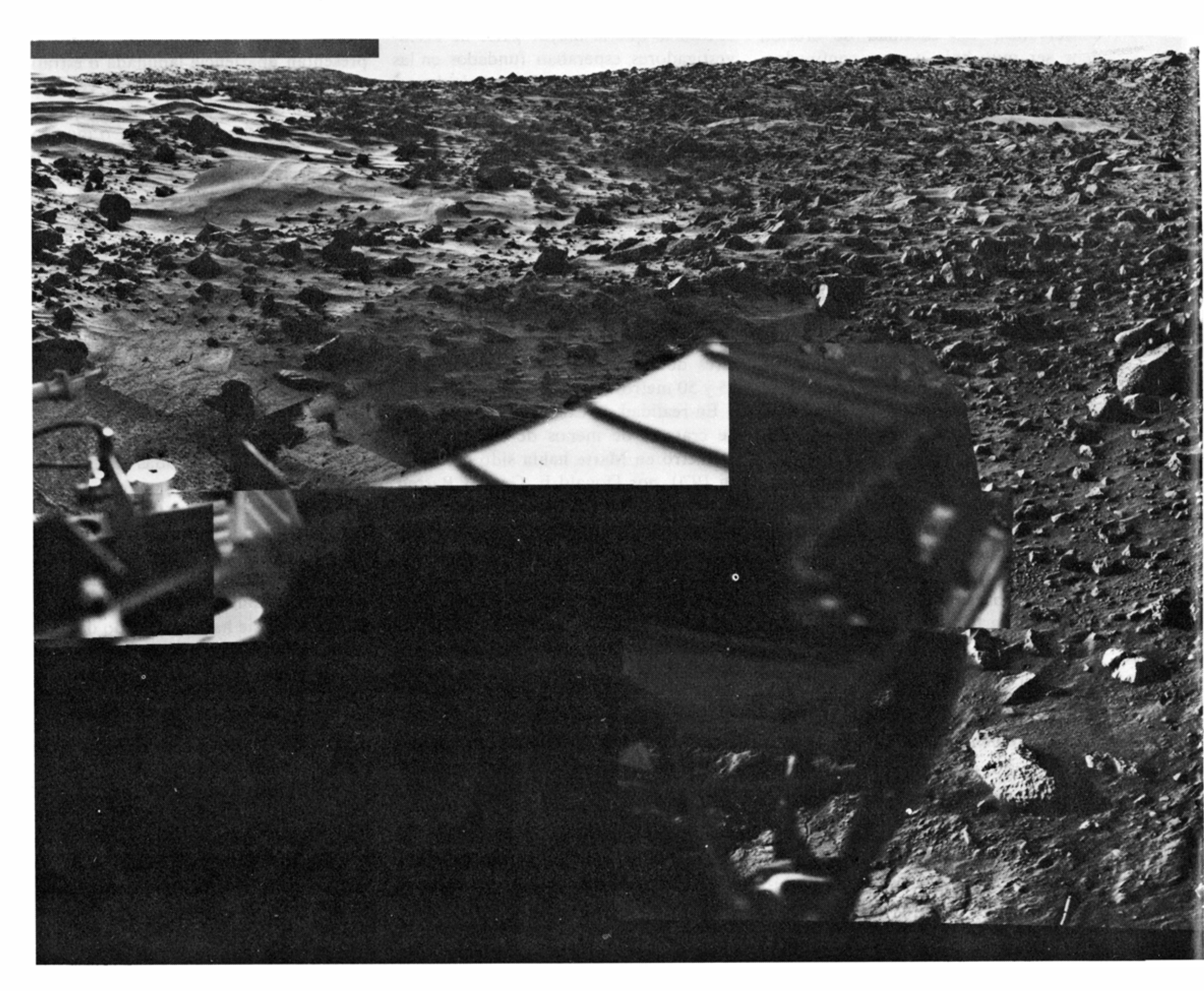
Otra característica del punto de descenso la reveló el descenso mismo. Cerca del módulo, el escape del retrocohete levantó materia suelta y dejó al descubierto una costra de suelo fracturada en forma poligonal. Tal suelo, conocido como duricostra, tiene un aspecto análogo al de los depósitos conocidos como "caliche" en el suroeste de los Estados Unidos y México. En la Tierra se forma la duricostra cuando una solución diluida de sal emigra a través del suelo hacia arriba; el agua se evapora de la solución

y la sal y otras substancias se reúnen inmediatamente debajo de la superficie. Probablemente, el mismo proceso ha actuado en Marte. Se ignora si allí el agua procede de los poros, relativamente grandes, bajo la superficie, o bien de delgadas capas de agua entre granos de materia. Lo último parece más probable, porque las variaciones en el contenido de vapor de agua de la baja atmósfera indican que el agua está sometida a un ciclo regular entre la atmósfera y el suelo.

La prueba directa de la existencia de sales en la duricostra marciana se obtuvo tomando una muestra del suelo delante del módulo y analizándola químicamente con el espectrómetro de fluorescencia de rayos X que llevaba el módulo. Priestley Toulmin III, del Servicio de Exploración Geológica de los Estados Unidos, y los miembros de su grupo de rayos X, hallaron que la cantidad de azufre –elemento que es muy probable encontrar combinado en sales minerales– era algo mayor en terrones de suelo que en suelo desmenuzado. Los terrones abundan alrededor del módulo y probablemente son trozos de duricostra fracturados por el descenso y también por el viento natural.

La superficie de Utopia Planitia

La búsqueda del punto de descenso del módulo *Viking 2* se basaba, en parte, en el deseo de los investigadores de que el segundo módulo se posara en una región claramente diferente de la del *Viking 1*, pero todavía lo suficientemente lisa como para que el descenso tuviera éxito. La preocupación fundamental, sin embargo, era encontrar un lugar con alto contenido de vapor de agua para hacer máxima la probabilidad de hallar pruebas de vida. La región elegida fue la superficie de Utopia Planitia, que es parte del manto de rocas desmenuzadas en las vastas llanuras del hemisferio norte. Las observaciones hechas desde la órbita mostraban que la superficie está cortada por fracturas que dividen el terreno en formas poligonales de varios kilómetros de lado. El punto de descenso (48 grados norte de latitud, 225,6 grados oeste de longitud) está a 200 kilómetros al suroeste del gran cráter Mie, de 100 kilómetros de diámetro.



CHRYSE PLANITIA, el punto de descenso del módulo *Viking 1*, es una llanura ondulada con bloques esparcidos sobre ella. Justamente visibles a la derecha, del centro, se encuentran varias zonas de roca a la intemperie. Bajo ellas, se aprecian depósitos lineales o regueros de sedimento que se extienden desde las rocas. Los regueros se extienden aproximadamente de norte a sur. Hacia la parte superior izquierda se distingue un gran campo de material arrastrado por el viento, donde los depósitos se acumularon durante un período en que los vientos también soplaban de norte a sur. A causa de la geometría panorámica de la imagen, parece que los regueros cambian de dirección con el azimut.

Poco tiempo después de haberse posado sobre la superficie el módulo de descenso *Viking 2,* quedó claro que, desde el suelo, el punto de asentamiento de este módulo era superficialmente análogo al de descenso del *Viking 1:* la superficie era una llanura de duricostra sembrada de rocas. Sin embargo, aquí acababa la semejanza. El módulo de descenso *Viking 2* se posó sobre una llanura nivelada, en que la mayor parte del relieve topográfico lo crean las trincheras que dividen el terreno en polígonos. No se pueden observar afloramientos de rocas. En lugar de ello, los guijarros y cantos rodados están o bien parcialmente embutidos en la base de material finamente granulado, o bien reposan encima de él. Aunque hay otras explicaciones posibles de la morfología del punto de descenso, es probable que el módulo de descenso *Viking 2* fuera a posarse sobre una de las extensiones de materia residual que fluyeron desde el cráter Mie. Las corrientes de materia residual de la Tierra suelen arrastrar consigo grandes rocas, dejando un campo de guijarros incrustados parcialmente en el material de grano más fino.

Las depresiones en forma de trinchera visibles desde el módulo de descenso *Viking 2* tienen una anchura aproximada de un metro y 10 centímetros de profundidad, y los bordes de los surcos son ligeramente protuberantes. Los surcos visibles desde el módulo de descenso son mucho menores que las trincheras que se distinguen desde la órbita, pero es probable que todos ellos tengan un origen análogo. El tamaño y la forma de los surcos que se ven desde el módulo de descenso los convierten en una buena analogía del "suelo con dibujo" que se encuentra en regiones frías de la Tierra. En ella, el suelo con dibujo se forma en suelos saturados de hielo donde las bajas temperaturas hacen que el terreno se contraiga y fracture; las fracturas cortan generalmente el suelo constituyendo un dibujo poligonal. En primavera, el suelo congelado se deshiela y las fracturas se llenan de agua. En otoño, ésta se hiela y el invierno siguiente, el hielo, que es más débil que el suelo helado, se fractura de la misma manera. Repetidos ciclos del proceso dan lugar a un terreno que está cortado por surcos formando un



El centro del mosaico apunta hacia el sureste. La totalidad del mosaico cubre 160 grados de azimut. La topografía local es tal que la distancia al horizonte varía desde decenas de metros, a la izquierda, a varios kilómetros, a la derecha. El terreno en la parte inferior se encuentra sólo a un par de metros de la cámara. La escala se puede juzgar por la espiga del equipo de toma de muestras de suelo, objeto cilíndrico brillante que descansa en el suelo en la parte media baja del mosaico; tiene 10 centímetros de largo. La pata del módulo se ve en la parte inferior derecha y otras partes del vehículo se observan a la izquierda. Bajo el vehículo se ve un suelo fracturado del tipo de una corteza llamado duricostra.

dibujo poligonal, con los surcos ocupados por crestas de hielo.

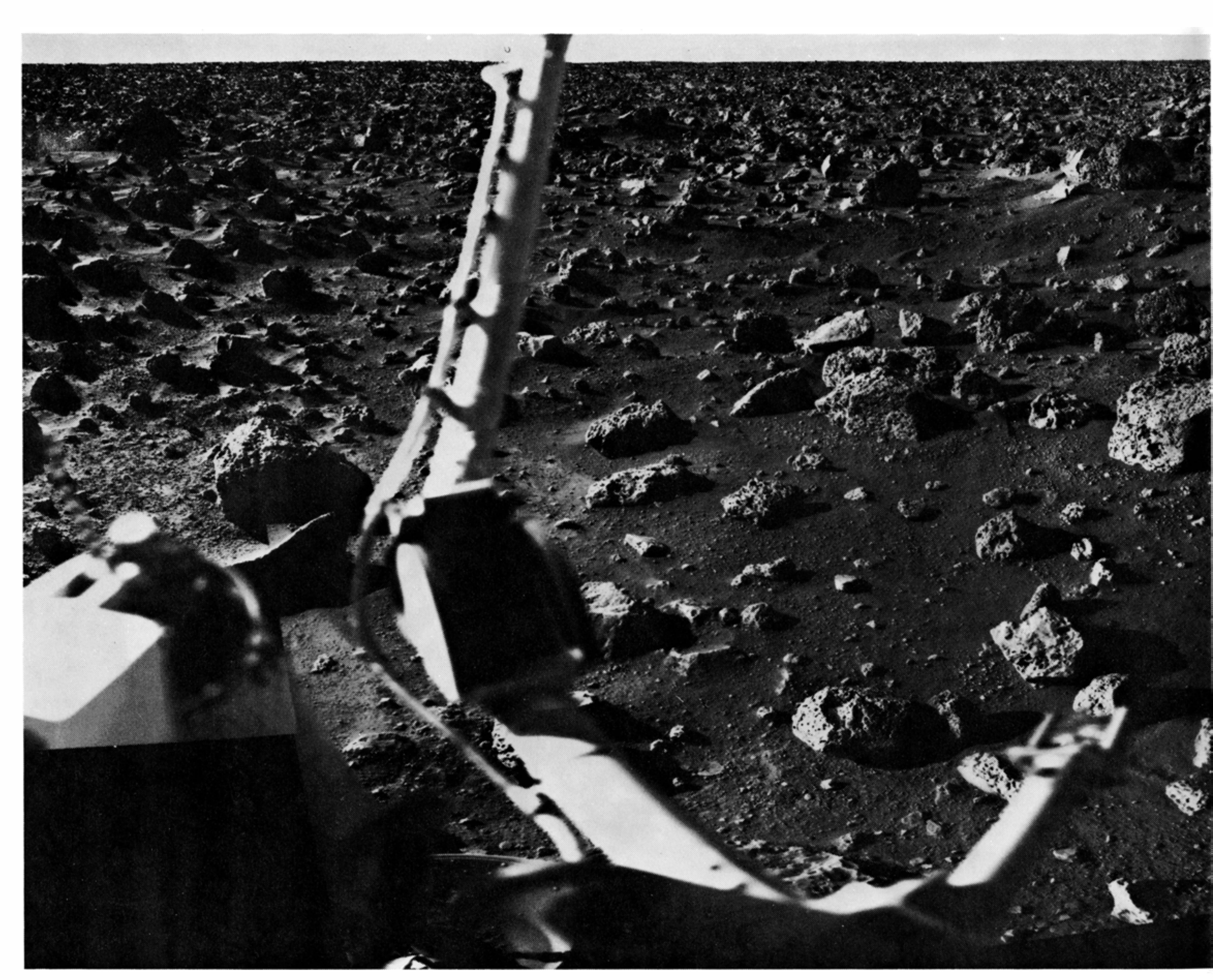
La idea de que este proceso sea el mismo que está actuando en Marte presenta la dificultad de que las temperaturas en Utopia Planitia siempre están bajo el punto de congelación del agua. Benton C. Clark, de la Compañía Martin Marietta, ha estimado que aunque el hielo contuviera sales, el punto de congelación de la disolución estaría sólo de 10 a 20 grados Kelvin por bajo del punto de congelación del agua pura. A las temperaturas de Utopia Planitia, el hielo impurificado continuaría sin fundirse.

Hay otro proceso que puede originar el suelo con dibujo: la desecación de arcillas. Las arcillas, que son el candidato más probable para el puesto de componente principal de los suelos marcianos recogidos en los puntos de descenso, se llegan a dilatar y contraer hasta en un 20 por ciento cuando absorben agua y luego la pierden. Si el terreno sobre el que descansa el módulo de descenso *Viking 2* estuvo en alguna época saturado de agua y después se secó, se pudieron haber formado grietas que originaron los surcos del dibujo poligonal.

Se pueden observar pequeñas vetas extendiéndose viento abajo desde algunas rocas en el punto de descenso del módulo *Viking 2,* lo mismo que se encontraban en el del *Viking 1.* Además, varios montoncitos ocupan el fondo de una vaguada grande delante del módulo. Una vez más, los montones indican que el viento dominante sopla aproximadamente de norte a sur; así, Chryse Planitia y Utopia Planitia pueden haber estado sometidas al mismo régimen de vientos. Como los puntos de descenso están separados por media vuelta al planeta, tal régimen de vientos tiene que haber sido global. Con toda probabilidad, tanto los amontonamientos como las vetas se formaron durante las tempestades de polvo que se producen cuando Marte está a su mínima distancia del Sol y las corrientes atmosféricas en gran escala en el hemisferio norte pueden estar dirigidas de norte a sur. Las estaciones meteorológicas de los módulos de descenso Viking, que registraron los vientos durante las tempestades de polvo de 1977, no mostraron, sin embargo, ningún predominio de las corrientes de norte a sur.

El suelo marciano

El espectrómetro de fluorescencia de rayos X de cada módulo de descenso analizó muestras de suelo y determinó



UTOPIA PLANITIA, punto de descenso del módulo *Viking 2,* se parece superficialmente a Chryse Planitia en que hay bloques esparcidos sobre el terreno. Este, sin embargo, es notablemente llano, y el horizonte se encuentra a varios kilómetros. El bloque mayor del centro de la imagen, que mide 35 centímetros de ancho y está a 2,75 metros del vehículo espacial, da una idea de la escala. No parece haber lecho rocoso a la intemperie. Una depresión en forma de surco claramente visible, de aproximadamente un metro de anchura y 10 centímetros de longi-

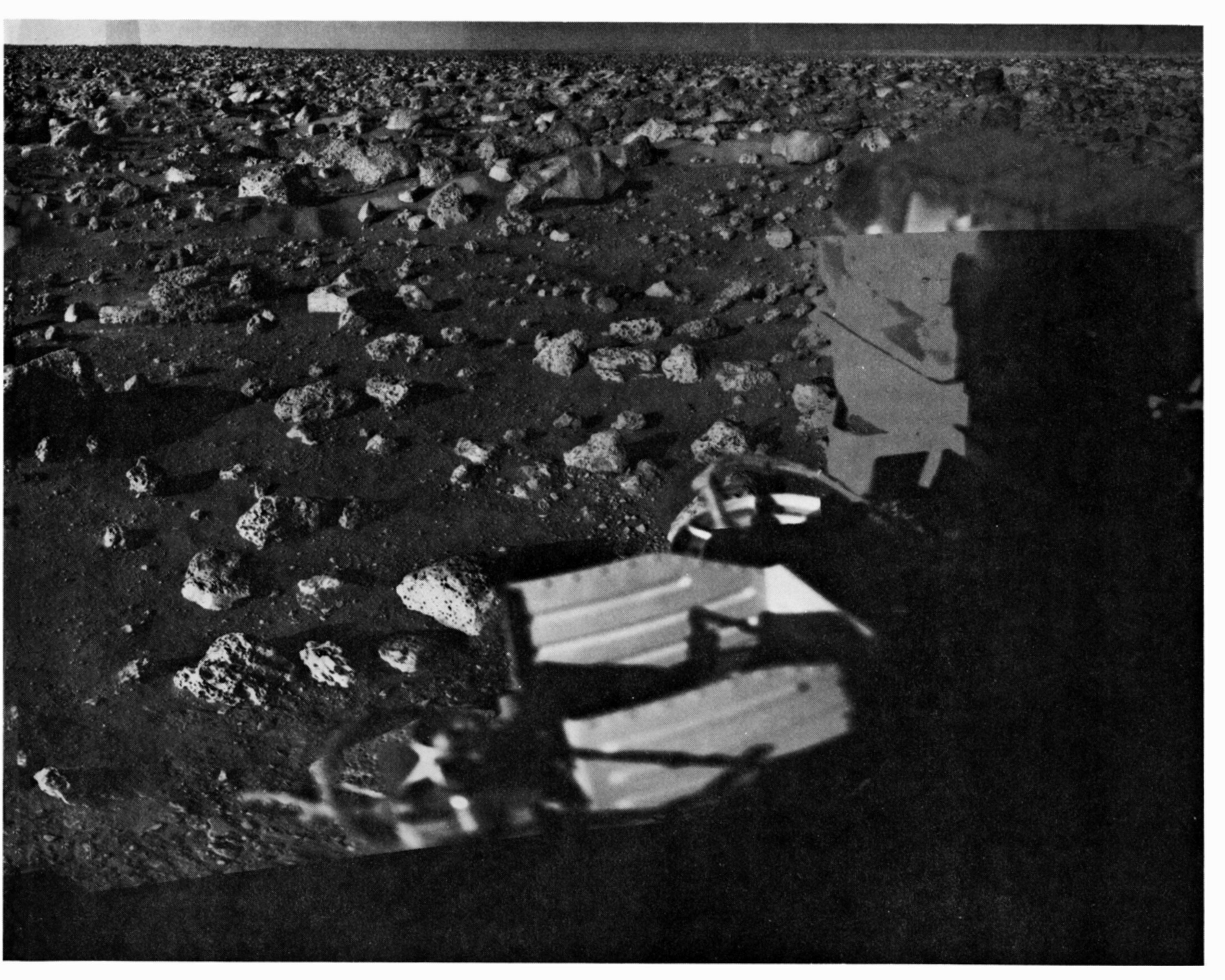
la abundancia de cierta cantidad de elementos con número atómico superior al del sodio (número atómico 11). Toulmin y colaboradores han examinado los datos y han mostrado que, en términos generales, la composición del suelo es aproximadamente la misma en ambos lugares. La composición difiere de la de cualquier mineral individual o tipo de roca conocido, lo que pone de manifiesto que el suelo es probablemente una mezcla complicada de materiales.

La superficie de Marte se halla constituida, al parecer, por terreno procedente de rocas ígneas máficas, es decir, rocas que han cristalizado a partir de una masa fundida, rica en hierro y magnesio. Comparadas con las rocas de la Tierra en general, las de Marte son ricas en magnesio, hierro y calcio y pobres en potasio, silicio y aluminio. Tales abundancias son compatibles con la clase de materiales que podría esperarse de una fusión parcial del manto de Marte, es decir, de la espesa capa que subyace a la corteza.

Es probable que el suelo analizado por los espectrómetros de rayos X de los Viking sea una mezcla de minerales de arcilla ricos en hierro, hidróxidos de hierro, minerales sulfatados y minerales carbonatados. Tal deducción concuerda con los resultados de los experimentos combinados de cromatografía de gases y espectrometría de masas realizados en cada módulo, los cuales mostraron que cuando se calentaban muestras de suelo, se desprendía vapor de agua y dióxido de carbono. El suelo contiene aproximadamente un uno por ciento de agua en peso, parte de la cual está probablemente ligada en minerales hidratados.

En la Tierra, el agua altera los materiales máficos y da lugar a suelos arcillosos ricos en hierro. El mismo tipo de proceso puede haber actuado en Marte en el pasado, cuando había agua líquida en abundancia. Parte del suelo se puede también haber formado al penetrar magmas incandescentes en la corteza y regolita cargadas de hielo, produciendo pequeñas erupciones que darían lugar a materiales esponjosos y arcillosos. Además, si hubiera habido agua en cantidad suficiente, el calor de los impactos podría haber bastado para transformar el material de rocas volcánicas en arcillas.

Robert L. Huguenin, de la Universidad de Massachusetts, ha propuesto una hipótesis diferente e intrigante. Sugiere que el suelo ha sido en gran parte producido por oxidación estimulada por irradiación ultravioleta de las rocas. La



tud, corta horizontalmente la parte media de la fotografía. Varios montones de material acumulado por el viento ocupan su parte baja. Hacia el extremo izquierdo, se ven depósitos lineales o regueros que se extienden viento abajo desde las rocas. Los regueros tienen aproximadamente el mismo azimut que los correspondientes de Chryse Planitia. La altiplanicie clara en el horizonte a la derecha está aproximadamente en la dirección del gran cráter Mie. Las temperaturas están bajo el punto de congelación del agua; el hielo impurificado no se fundiría.

radiación ultravioleta del Sol no se absorbe en la atmósfera marciana porque ésta no tiene ozonosfera. En presencia de pequeñas cantidades de vapor de agua, tal radiación es capaz de romper la estructura de los minerales de aluminio y silicio provocando que iones como los de hierro emigren hacia la superficie, con lo que se altera la estructura cristalina del mineral. No se conoce hasta qué punto el suelo se puede haber formado mediante tal proceso.

En la azada que formaba parte de la pala de recogida de suelo en cada módulo de descenso, se montaron dos imanes. Uno de ellos estaba a nivel de la superficie de la azada; el otro estaba embutido en el metal, de manera que su fuerza efectiva era de una doceava parte de la del primero. En ambos lugares, después que la azada se había hundido en el suelo, quedaron adheridas a ambos imanes cantidades iguales de materia. Según Robert B. Hargraves, de la Universidad de Princeton, y David W. Collinson, de la Universidad de Newcastle upon Tyne, para que quedaran adheridas a ambos imanes porciones iguales de material, éste tiene que haber contenido una proporción de entre 3 y 7 por ciento de materia magnética. Si el suelo hubiera contenido una concentración más baja de materia magnética, las porciones no podrían haber quedado adheridas al imán más débil. Por tanto, existe material magnético en cantidades apreciables en el suelo marciano.

El color del material que quedó adherido a los imanes es igual que el de la superficie, lo que indica que la materia magnética está cubierta o teñida por los mismos materiales que colorean el suelo marciano. Parece razonable suponer que los minerales magnéticos sean magnetitas o maghemitas, que son ambos óxidos de hierro, y níquel-hierro metálicos. La maghemita es de color pardo amarillento o rojizo, y, si se encuentra presente, puede contribuir al color de Marte. La abundancia de material magnético y su naturaleza probable concuerda bien con la hipótesis de una procedencia máfica, tal como los basaltos máficos.

Las cámaras instaladas en los módulos de descenso Viking sacaron fotografías de la superficie y el cielo en seis regiones del espectro, extendiéndose desde unos 0,5 micrometros (en el azul) hasta 1,0 micrometros (en el infrarrojo próximo. Las fotografías en color se produjeron determinando primero la irradiancia espectral de Marte en cada una de las regiones y calculando entonces el tono, brillo y saturación de color para el intervalo de longitudes de onda al que es sensible el ojo humano. La irradiancia espectral de la superficie de Marte es un producto de la irradiancia del sol, la reflectancia de la superficie de Marte y la naturaleza de la luz difundida por la atmósfera. Las fotografías en color muestran una superficie parda amarillenta. Si se eliminan los efectos del color introducido por la atmósfera, la superficie tiende a ser todavía más parda.

Los espectros de reflectancia deducidos de los datos de las cámaras para varias zonas de suelo en ambos puntos de descenso son notablemente similares y parecen variar solamente cuando lo hace la iluminación. Los espectros tienen un aspecto semejante al de los obtenidos en las áreas claras de Marte con instrumentos situados en la Tierra. Se cree que tales áreas están cubiertas por suelo alterado químicamente y de grano fino. Una explicación que se ajusta a los datos disponibles es que la mayor parte del terreno a la intemperie que se ve en los puntos de descenso y se ha recogido para análisis, está constituido por un producto de color claro resultante de erosión química que el viento ha mezclado uniformemente en todo el planeta.

La única área observada en los puntos de descenso cuyo espectro de reflectancia es apreciablemente diferente del de todas las otras zonas es un reguero de materia a unos 15 metros del módulo de descenso *Viking 1*. Este reguero, más obscuro que el suelo que lo rodea, puede haberse quedado atrás después que el viento arrastró el material más claro, que era también más fácilmente movible. El espectro del reguero obscuro es semejante al obtenido con instrumentos situados en la Tierra para las regiones obscuras de Marte. El reguero puede estar constituido por rocas ígneas ricas en hierro y parcialmente alteradas, que se han desmenuzado hasta el tamaño de partículas del suelo. Así, aunque en general hay una marcada similitud entre los suelos en ambos puntos de descenso, existen diferencias que pueden imitar, en pequeña escala, las diferencias entre las clásicas áreas claras y obscuras de Marte que se ven desde la Tierra.

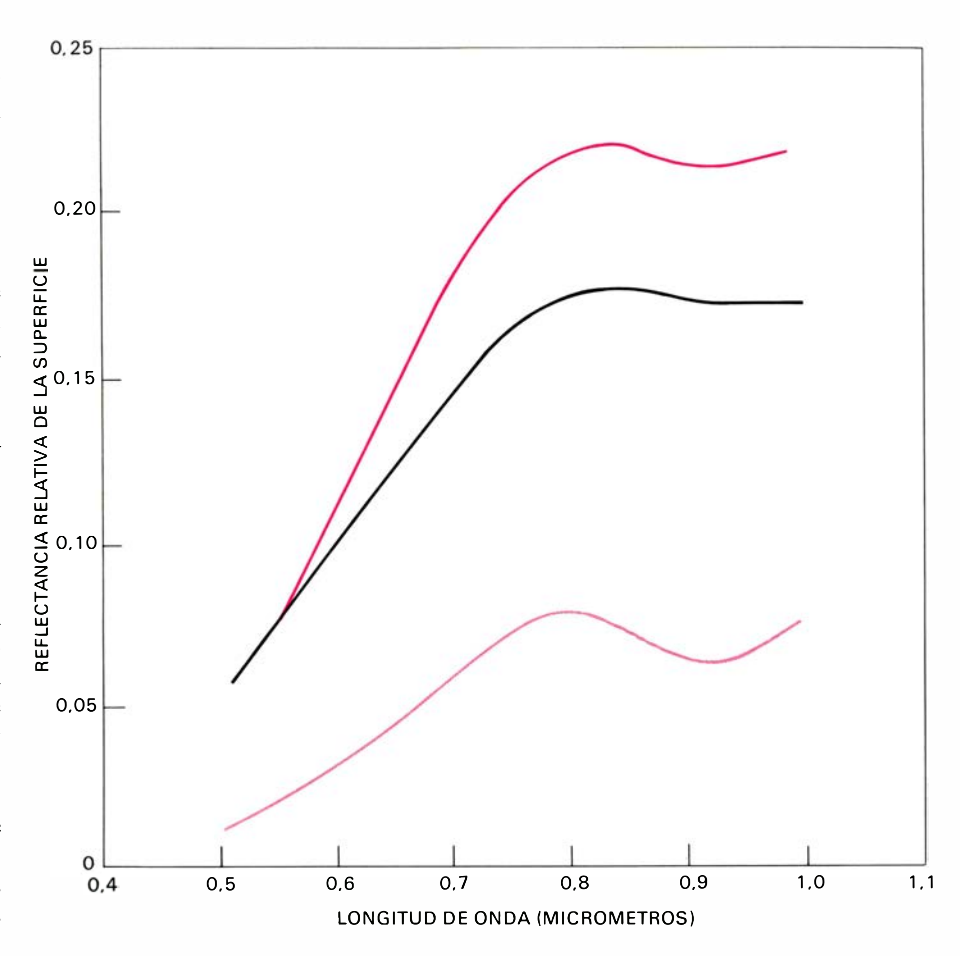
La mayor parte de las rocas y los afloramientos del lecho rocoso que se ven en las fotografías en color, tomadas por los módulos de descenso, aparecen más obscuras que el suelo sobre el que reposan, pero cuando se observan bajo iluminación comparable, la mayoría resultan en realidad más claras. En la mayor par-



DOS MONTONES de material de grano fino se encuentran a 15 metros del módulo de descenso *Viking 1*. El montón de la derecha está formado de material más obscuro que el situado a la izquierda. Los espectros del montón obscuro indican que es el único ejemplo de tal material hallado en cualquiera de los dos puntos de descenso. Todas las demás zonas de suelo que se distinguen desde los módulos son muy similares. Resulta bastante probable que pertenezcan a la misma clase que los suelos que predominan en las regiones claras del planeta Marte.

te de las fotografías, el ángulo entre el Sol, la superficie y la cámara es mayor para los lados de las rocas que para la superficie del suelo, y, por ello, la claridad de las rocas aparece menor. Las rocas y el suelo tienen formas espectrales semejantes, lo cual pone de manifiesto que las rocas están cubiertas de una suave capa de recubrimiento o colorante de materia cuya composición es análoga a la del suelo. El aspecto general de la mayoría de las rocas sugiere un origen volcánico de las mismas, aunque la erosión del viento haya hecho apreciable mella en algunas. La mayoría de las rocas del punto de descenso del *Viking 2* y algunas de las halladas en el del *Viking 1* presentan oquedades como las que se forman en la Tierra a partir de lavas ricas en gases. Los gases disueltos en la lava, a causa de las altas presiones a gran profundidad, se desprenden cuando la lava alcanza la superficie y forman bolsas de gas. Cuando la roca fundida se enfría y endurece, las bolsas se conservan en forma de burbujas o vesículas en la roca. Por desgracia, resulta extremadamente difícil distinguir entre las oquedades producidas por este proceso y las causadas por erosión del viento sobre minerales más blandos.

Uno de los aspectos mas intrigantes de ambos puntos de descenso, y probablemente de Marte en general, es el hecho de que parece haber un marcado déficit de arena comparado con un desierto típico de la Tierra. La falta de arena en los puntos de descenso ha sido deducida por Henry J. Moore II, del Servicio de Exploración Geológica de los Estados Unidos, y sus colaboradores a partir de las fotografías de las paredes de las trincheras excavadas por el dispositivo de toma de muestra del suelo en los dos módulos de descenso Viking. Sus resultados indican que la parte principal de las partículas pequeñas de Marte no alcanza los 0,1 milímetros. Tales partículas serían difíciles de arrancar de la superficie, y una vez arrancadas serían probablemente arrastradas por el viento.



ESPECTROS DE REFLECTANCIA de los montones claro y obscuro en el punto de descenso del módulo *Viking 1*, junto con el de una trinchera excavada allí por el dispositivo de toma de muestras del suelo. Los espectros se obtuvieron en el Centro de Investigación Langley de la Administración Nacional de Aeronáutica y del Espacio (NASA), mediante una técnica especializada ideada por Friedrich O. Huck y Stephen K. Park. Los espectros muestran la fracción de radiación solar recibida que se refleja para longitudes de onda entre 0,5 micrometros (en el azul) y 1,0 micrometros (en el infrarrojo próximo). Las formas de los espectros para el montón claro *(curva obscura)* y para la trinchera *(curva clara)* son similares. Ambos muestran una banda de absorción a 0,93 micrometros, lo que implica que tienen análoga composición. La única diferencia importante entre ellos es la intensidad de la luz reflejada. (La diferencia era de esperar, en realidad, porque el suelo de la trinchera había sido alterado al excavarla, aumentando en consecuencia la microtopografía del suelo e incrementando su grado de dispersión y sombreado y disminuyendo su reflectancia.) El espectro de reflectancia del montón obscuro *(curva negra)*, sin embargo, tiene una forma diferente. Ha desaparecido la banda de absorción a 0,93 micrometros, dando al espectro una forma aplanada, y puede haber una banda débil cerca de 1,0 micrometros. La diferencia entre los montones claros y obscuros puede imitar la diferencia entre las grandes regiones claras y las obscuras de Marte tal como se contemplan desde la Tierra. Ilustración diseñada por Ilil Arbel.

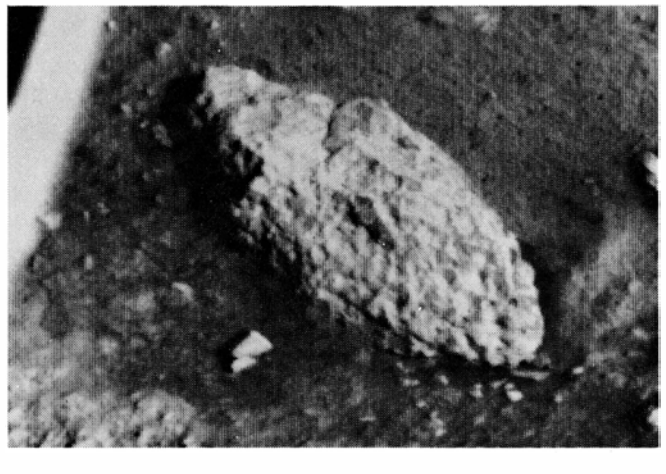
Los vientos de Marte

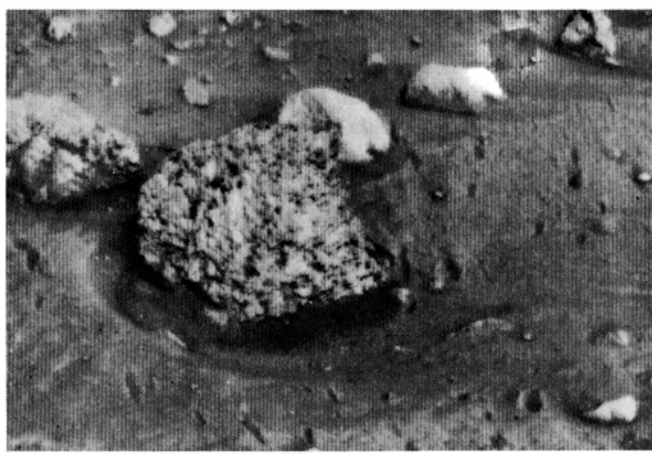
Probablemente, en Marte hay partículas de tamaño de las de arena en las dunas que rodean los casquetes polares marcianos; quizás incluso constituyen los regueros obscuros que se aprecian en el punto de descenso del *Viking 1* y en porciones de las áreas obscuras clásicas que cubren las latitudes medias del hemisferio sur. El suelo marciano, sin embargo, no contiene la clase de arena que se encuentra en la Tierra. La mayor parte de la arena terrestre está constituida por minerales de cuarzo y feldespatos que se han formado por la acción del tiempo y clima a partir de rocas ácidas ígneas y metamórficas. El cuarzo y el feldespato, que dominan las rocas ácidas, son resistentes a la modificación química y al ataque mecánico, y los dos minerales son componentes importantes de las rocas sedimentarias de la Tierra. Marte no ha evolucionado hasta el punto de haber creado rocas ácidas en grandes cantidades; el planeta está probablemente dominado por basaltos máficos. En tales basaltos, los minerales más importantes son el olivino, piroxeno y la plagioclasa, feldespatos todos ellos. Huguenin ha mostrado que incluso en el frío y árido ambiente de Marte, esos minerales son modificados rápidamente por oxidación estimulada por los rayos ultravioleta.

Si, como se sospecha, el suelo claro de los lugares de descenso está formado por minerales arcillosos, el grano que lo forma sería muy fino. Podrían existir, sin embargo, partículas del tamaño de granos de arena como conglomerados de otras partículas más pequeñas. Tales conglomerados del tamaño de la arena tendrían una vida muy corta en el sistema eólico marciano. Los granos que rebotan en la superficie viajan, aproximadamente, con la velocidad del viento. Como la atmósfera de Marte tiene una densidad aproximada de sólo la centésima parte de la densidad de la Tierra, se necesitarían vientos de unos 50 metros por segundo (180 kilómetros por hora)

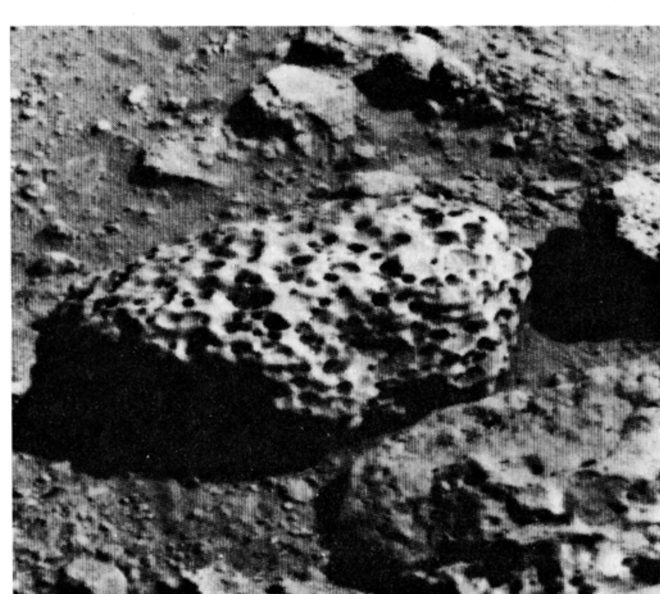
para arrancar partículas del tamaño de granos de arena de la superficie marciana. Las partículas de arena terrestre transportadas a esa velocidad serían poderosamente abrasivas, pero los conglomerados de material arcilloso se desintegrarían al chocar, convirtiéndose en inofensivas motas de polvo. Incluso granos minerales tales como olivino, piroxeno o feldespato serían fáciles de pulverizar si fueran transportados a esa velocidad. Una consecuencia de esta situación en Marte es que la proporción de erosión eólica resulta pequeña comparada con la de la Tierra. La baja velocidad de erosión probablemente explica por qué gran parte de la superficie de Marte tiene un aspecto tan afilado y primitivo. El material suelto se transporta fácilmente, pero las rocas se erosionan con gran lentitud.

Los cielos, en los puntos de descenso, son pardo-amarillentos y se han mantenido de ese color a lo largo del año marciano que los módulos de descenso han estado en Marte. Esta observación es algo inesperada. De acuerdo con cálculos de Pollack, el color se debe principalmente a partículas de polvo suspendidas en la atmósfera hasta unos 40 kilómetros de altura. Era de esperar que quedaran suspendidas en la atmósfera grandes cantidades de polvo tras las tempestades de polvo importantes, que tienen lugar cuando el planeta pasa por el perihelio, es decir, el punto de su órbita más próximo al Sol. Las fotografías de los módulos de descenso mostraron, sin embargo, que los cielos eran pardo-amarillentos aun cuando el planeta se encontrara en el afelio, punto de su órbita más alejado del Sol. En esa época son raras las tempestades de polvo importantes. O bien el polvo es elevado frecuentemente en Marte, o bien la turbulencia atmosférica lo sostiene dinámicamente durante largos períodos de tiempo. No obstante, incluso en los días de mayor contenido de polvo, si todo éste se depositase sobre la superficie, formaría una capa de sólo una fracción de milímetro de espesor.









ESTOS PRIMEROS PLANOS de rocas marcianas ilustran la variedad de rocas en el planeta. Las dos fotografías de arriba muestran rocas de Chryse Planitia; las dos de abajo son de Utopia Planitia. La roca de la parte superior izquierda tiene unos 25 centímetros de largo; su aspecto picado y variopinto recuerda las rocas ígneas de la Tierra que han sufrido erosión por parte del viento. En la parte inferior izquierda puede verse la duricostra dejada al descubierto por el escape del retrocohete. La roca en la parte superior derecha parece ser una roca volcánica del tipo brecha, de unos 20 centímetros de ancho. A un lado de la roca puede verse una oquedad excavada por el viento; desde el otro lado se extiende un depósito prominente de material depositado por el viento. Las oquedades alargadas en el suelo se produjeron al chocar con éste trozos que habían sido proyectados por el descenso del módulo. La roca en la parte inferior izquierda es un bloque rectangular de unos 40 centímetros de ancho. La forma angular de la roca indica que en su aspecto dominan planos de fractura que forman ángulos rectos entre sí. Al fondo se ve material arrastrado por el viento, así como una roca cónica cuya forma puede deberse a erosión eólica. La roca en forma de cacahuete de la fotografía inferior derecha puede ser un trozo de lava de unos 30 centímetros de ancho. Las oquedades que presenta se pudieron formar cuando gas disuelto en la lava originó burbujas o vesículas y después se desprendió, o bien pueden haber sido porciones de minerales más blandos erosionados por el viento. Los puntos de descenso exhiben gran variedad de formas de rocas, unas erosionadas, otras de aspecto primitivo.

Los módulos de descenso han recogido datos durante dos grandes tempestades de polvo, ambas iniciadas en el hemisferio sur de Marte cuando el planeta se acercaba al perihelio en 1977. La primera tempestad comenzó en febrero y la segunda en mayo. La nube de polvo fue distribuida rápidamente alrededor de todo el planeta por los vientos a gran altura. Pasaron en ambos casos varios meses antes que la tempestad amainara. En los puntos de descenso en el hemisferio norte, el viento sólo en raras ocasiones alcanzó velocidad suficiente para afectar a la superficie del suelo. Parece probable que tras ambas tempestades se acumulara una fina capa de polvo sobre los módulos y las superficies alrededor de ellos. La capa de polvo indica que no toda la materia alzada del hemisferio norte es devuelta a éste. En un período de tiempo suficientemente largo, las tempestades de polvo deberían desproveer de polvo las latitudes medias del hemisferio sur, exponiendo el lecho rocoso y dejando tras sí depósitos de materia más obscura y menos afectada por la intemperie. Quizá tales depósitos son las clásicas zonas obscuras de Marte que se ven desde la Tierra.

La latitud de Marte que está directamente bajo el Sol en el perihelio es el punto donde comienzan la mayoría de las tempestades de polvo. Este punto subsolar en el perihelio varía lentamente con el tiempo porque el eje de rotación de Marte presenta precesión. La precesión marciana debe producir una migración de dicho punto subsolar entre +25 y –25 grados de latitud en un período de 50.000 años. Lawrence A. Soderblom, del Servicio de Exploración Geológica de los Estados Unidos, ha indicado que si las zonas obscuras de Marte son regiones despojadas de una fracción relativamente grande de la capa clara de polvo móvil, podrían oscilar en torno al ecuador con el mismo período. Dicho de otra manera, hace unos 20.000 años, la mayor parte de las tempestades de polvo pudieron haber comenzado a la latitud que ahora ocupa el módulo de descenso *Viking 1*, y Chryse Planitia pudo haber sido parcialmente despojada de sus depósitos claros. Si éste es el caso, los regueros y amontonamientos que se ven en el punto de descenso del *Viking 1* quizá tengan menos de 20.000 años.

El futuro

No cabe duda de que la misión Viking ha aumentado considerablemente nuestros conocimientos de la geología de Marte. Sabemos bastante acerca del aspecto que presenta la superficie y de la clase de materiales que contiene. El descubrimiento de que al menos uno de los casquetes polares residuales está constituido por verdadero hielo ha incrementado apreciablemente nuestro conocimiento respecto a la intensidad de las fluctuaciones climáticas del planeta. Quedan importantes problemas sobre la evolución de la superficie marciana. No se conocen con mucha exactitud las edades de las distintas clases de terreno del planeta. La estructura y composición del interior de Marte continúa siendo en gran parte un misterio. Sin tal conocimiento es imposible construir un modelo teórico único sobre la formación y evolución del planeta.

Algunas de las preguntas podrán responderse tras ulteriores análisis de los datos de los Viking. Otras tendrán que esperar misiones futuras. Una posible misión es un vehículo orbital, mucho más complicado que el satélite terrestre Landsat, capaz de cartografiar características químicas y mineralógicas de la superficie marciana. Otra posibilidad sería un vehículo de superficie no tripulado capaz de recorrer cientos de kilómetros a lo largo de varios años, que pudiese analizar la regolita con más detalle del que puede suministrar un satélite en órbita. Una tercera posibilidad sería una serie de cohetes penetrantes, lanzados desde una plataforma en órbita, que pudieran incrustarse en la superficie del planeta en varios puntos y constituir una red de sensores meteorológicos y sismológicos. Una cuarta posibilidad sería una misión que recogiese y devolviera a la Tierra muestras de material marciano, permitiendo obtener la clase de datos que sólo un laboratorio terrestre puede suministrar. Basta considerar la enorme cantidad de datos que se obtuvieron analizando las muestras traídas de la Luna por las misiones Apolo para apreciar la cantidad de información que se podría conseguir con muestras traídas de Marte.

La exploración futura de Marte se puede justificar desde varios puntos de vista. Quizás el más importante es que la Tierra y Marte parecen haber seguido caminos evolutivos lo suficientemente diferentes como para que cada uno de ellos haya tenido una historia única. Sus caminos de evolución, sin embargo, han sido lo suficientemente parecidos como para que tenga sentido la comparación de datos sobre sus atmósferas, superficie e interior. Pocas dudas pueden existir de que la comprensión de la historia de Marte incrementará nuestra comprensión de la Tierra.

Ciencia y sociedad

Teléfono electrónico

El objetivo básico de la telefonía es facilitar una comunicación oral a larga distancia. El enfoque seguido tradicionalmente ha sido convertir las ondas sonoras producidas por la voz del usuario en energía eléctrica que pueda transmitirse de un modo eficaz por hilos conductores. Las ondas eléctricas, teóricamente idénticas en forma y frecuencia a las ondas sonoras originales, al llegar al auricular del microteléfono del receptor se convierten de nuevo en ondas sonoras que corresponden, más o menos, con el modelo de habla emisor.

Adicionalmente, y a fin de poder establecer una comunicación, el aparato telefónico debe ofrecer al usuario la posibilidad de seleccionar cualquier otro aparato del sistema, e indicarle si éste se encuentra ocupado. Ningún teléfono puede realizar estas funciones sin el soporte de un sistema de interconexión. Cada teléfono perteneciente a un área geográfica determinada está conectado (por medio de un par de hilos de cobre llamado bucle o línea) a una central telefónica que opera como centro de conmutación. Las llamadas telefónicas dentro de la misma área se dirigen a través de la central, mientras que las llamadas a números fuera del área se dirigen por medio de enlaces entre centrales.

Imaginemos que un usuario se dispone a realizar una llamada telefónica. Una vez descolgado el microteléfono (aparato *A*), la central telefónica le envía el tono de invitación a marcar, haciéndole saber de este modo que la línea está libre. Seguidamente, marca un número telefónico (el número del aparato *B*); la central telefónica selecciona el par de hilos correcto y envía una corriente de llamada hacia el aparato *B*. Si alguien que esté en el aparato *B* descuelga el microteléfono y contesta, la central telefónica conecta *A* con *B*, estableciéndose así la comunicación. La central telefónica detecta también el final de la comunicación y libera las líneas para otras llamadas. Cuando se realiza una llamada de larga distancia, la central telefónica automáticamente registra la información necesaria para la tarificación. Las secciones que realizan estas funciones se denominan circuitos de supervisión.

El emisor es un micrófono controlado por el sonido que recibe. Funciona de acuerdo con el principio de que la vibración de un diafragma puede variar la tensión de una corriente eléctrica suministrada externamente. Ello se debe a la variación de resistencia de una pequeña cápsula, rellena de granalla de carbón. Cuando el diafragma se desplaza, un pequeño émbolo solidario a la misma varía la presión aplicada a la granalla de carbón, de forma que la resistencia total presentada por la cápsula experimenta un cambio aproximadamente igual a las variaciones del sonido incidente. Si se aplica un voltaje a la cápsula, a fin de producir una corriente eléctrica constante, las ondas de sonido que inciden sobre el diafragma producirán una modulación de la corriente. (La alimentación del teléfono se suministra desde la central telefónica por un sistema de baterías soportado por generadores auxiliares. Esta es la razón de que, por lo general, el servicio telefónico se mantenga cuando se produce una interrupción en la red de energía eléctrica.)

El receptor es un auricular formado por un diafragma y un imán permanente con una bobina arrollada al mismo. La señal eléctrica de llegada se aplica al devanado de la bobina, produciendo consecuentemente una perturbación en el campo magnético permanente. Tal perturbación hace vibrar el diafragma, el cual genera ondas sonoras similares a las de la conversación original.

El emisor y el receptor se conectan con el bucle a través del circuito de conversación, frecuentemente denominado red de paso de dos a cuatro hilos (bobina híbrida), debido a que los cuatro hilos del microteléfono se reducen a dos en la interconexión con el bucle. La función principal del circuito de conversación es separar los circuitos de habla y escucha a fin de limitar la fracción de la señal a transmitir (tono local) que se recoge en el auricular propio. El nivel de tono local es importante desde el punto de vista de las reacciones humanas. Un nivel de tono local elevado fatiga el oído y hace que el usuario baje el volumen de su voz, reduciéndose así el nivel de sonido que llega a su interlocutor. Asimismo con un nivel de tono local elevado, el ruido de fondo se reproduce en el receptor propio, degradándose por tanto la inteligibilidad de la voz recibida. Un nivel de tono local excesivamente bajo caracteriza al teléfono con un sonido poco natural o "muerto", y hace que el usuario tienda a hablar demasiado alto.

Otro factor que ha de tenerse en cuenta es la situación del teléfono con respecto a la central telefónica. La resistencia efectiva del hilo de cobre que constituye el bucle aumenta proporcionalmente con su longitud, atenuando en consecuencia las señales presentes en línea. Un problema adicional es la reducción de la eficacia del emisor de carbón conforme disminuye la corriente de alimentación al aumentar la resistencia del bucle.

El aparato telefónico convencional se ha diseñado para compensar en cierta medida tales pérdidas. Esta compensación se denomina ecualización, y se logra con la adición de dos varistores (resistencias variables), a la red de conversación. Estos componentes ofrecen una resistencia elevada para valores bajos de corriente continua, y viceversa. Por ser constante la tensión de alimentación que suministra la central al teléfono, la misión de los varistores es compensar el valor relativamente alto de corriente que se presenta en bucles cortos, que tiende a producir señales de voz excesivamente elevadas, y el valor relativamente bajo de corriente en bucles largos que tiende a hacer que el usuario alce su voz.

El timbre de un teléfono es un circuito electromagnético convencional que gobierna las percusiones de un badajo sobre dos campanas según una determinada secuencia de actuación. El gancho conmutador sirve para desconectar la red de conversación y el disco marcador, cuando el aparato no se utiliza. Su apertura se produce por el peso del microteléfono descansando sobre su soporte. Cuando se descuelga el microteléfono, se conecta el circuito de conversación al bucle, y se establece una circulación de corriente continua. Esta corriente excita un relé en la central telefónica que conecta el bucle correspondiente a los circuitos de supervisión. Seguidamente, la central envía hacia el teléfono un tono audible, indicativo para el usuario de que el sistema se encuentra preparado para recibir el número de la persona con la que desea comunicarse.

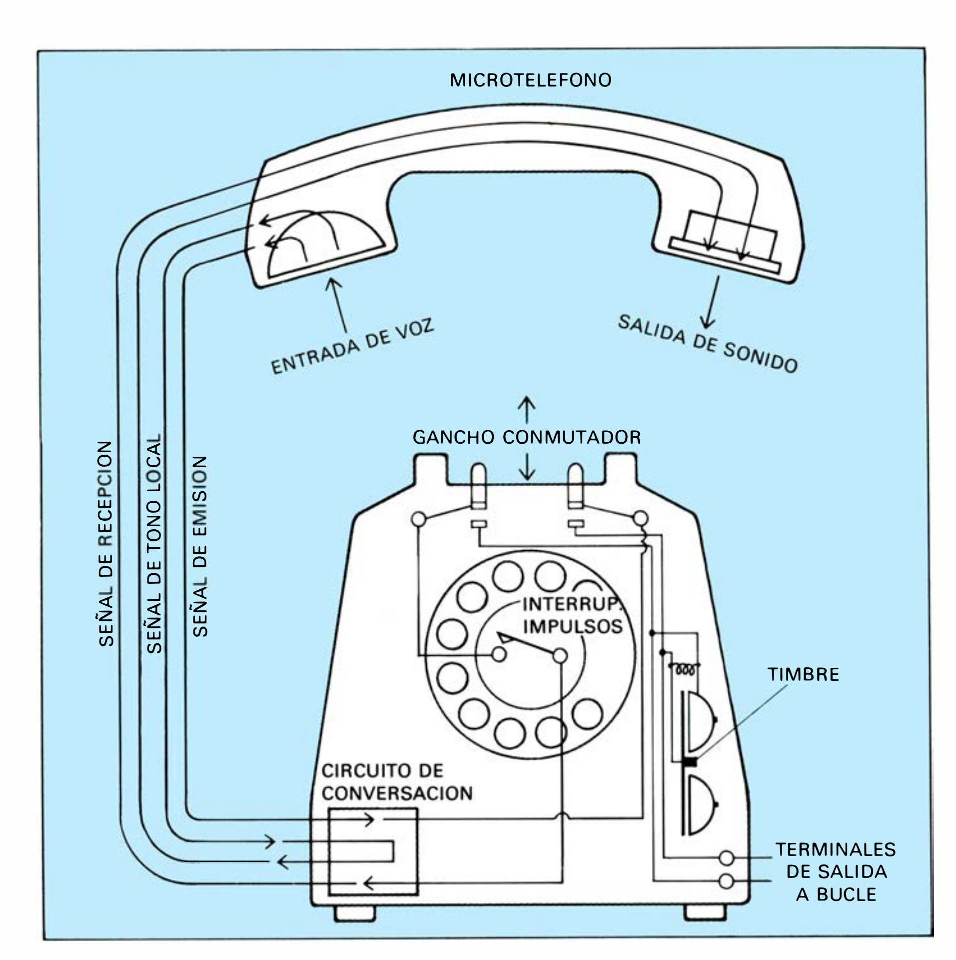
El método tradicional de marcar un número telefónico es por medio del disco. Si el usuario hace girar, por ejemplo, el orificio número 6 hasta el tope final, se "carga" un resorte y se excitan varios contactos. Cuando se libera el disco, la velocidad de retroceso se controla por medio de un regulador, de manera que só-

lo se envían hacia la central telefónica, con una velocidad correcta, seis impulsos de apertura. La velocidad más frecuente es 10 impulsos por segundo, con una interrupción aproximada de 1/16 de segundo entre un impulso y el siguiente. El dispositivo de marcación por teclado opera de manera diferente. Cada vez que se pulsa una tecla, se pone en funcionamiento un oscilador a transistor que genera un tono (diferente combinación de dos frecuencias seleccionadas entre 8 posibles valores). En la central telefónica un receptor-decodificador detecta y reconvierte los tonos.

Salvo en el sistema de marcación, no se ha operado ningún cambio ni en el diseño básico ni en las características funcionales. La aparición de un teléfono electrónico, que la firma Bell Canadá está sometiendo a pruebas de campo, puede considerarse un paso hacia adelante muy significativo. En este teléfono, los componentes electromecánicos convencionales, tales como el timbre, la bobina híbrida del circuito de conversación y el circuito oscilante del generador de tonos de marcación han sido sustituidos por circuitos integrados, uno para cada función. (Los circuitos integrados hacen posible la interconexión de cientos de transistores, diodos y resistencias sobre un único sustrato de silicio, con 96 milímetros cuadrados de dimensiones típicas). Los tres circuitos integrados están interconectados con ciertos componentes externos, conformando así el teléfono básico. Transductores electrodinámicos sustituyen al micrófono de carbón y a la cápsula receptora electromagnética.

El transductor electrodinámico se basa, para su funcionamiento, en el principio de que todo conductor eléctrico, que esté vibrando en el seno de un campo magnético, genera una diferencia de potencial entre sus terminales. El transductor electrodinámico consta de un fino diafragma al cual se ha fijado rígidamente una pequeña bobina. La acción de las ondas sonoras sobre el diafragma mueve la bobina dentro del campo radial del imán permanente, generando así una señal de corriente variable. Aunque la unidad electrodinámica se alimenta por sonido (mientras que el micrófono de carbón requiere una corriente eléctrica), la señal de salida es pequeña, y debe incorporarse un circuito amplificador para obtener los niveles adecuados de señal.

El circuito de conversación se ha diseñado para operar con transductores electrodinámicos, incluyendo, en consecuencia, amplificadores para ambos transductores, uno que actúa como micrófono y otro idéntico que se utiliza como auricular. La ganancia de los amplificadores se ajusta automáticamente, conforme varía la corriente entre bucles cortos y largos. De esta manera se igualan los niveles de señal para compensar las variaciones de longitud de unos bucles a otros. Un circuito de equilibrado facilita el nivel correcto de tono local. Las funciones de marcación se realizan en otro circuito adicional: el generador de tonos multifrecuencia. La marcación por teclado multifrecuencia se basa en una norma aceptada internacionalmente que comprende cuatro frecuencias pertenecientes al grupo bajo y tres frecuencias pertenecientes al grupo alto. (Una cuarta frecuencia perteneciente al grupo alto se reserva para que, en el futuro, los teclados puedan tener una cuarta columna, de manera que dispondrán 16 teclas en lugar de 12 como en la actualidad.)



Funcionamiento de un teléfono convencional (aquí, el modelo 500)

Supóngase que un usuario marca la tecla 5. La frecuencia correspondiente al grupo alto es de 1336 hertzios y la del grupo bajo 770 hertzios. Las frecuencias deben tener una tolerancia de 1,5 por ciento con respecto a su valor nominal bajo todas las condiciones de temperatura durante el período de vida del equipo. El circuito consigue este nivel de precisión utilizando un método de división digital. El oscilador de alta frecuencia generada por un oscilador, que actúa como reloj maestro, se divide sucesivamente en dos contadores, uno para cada grupo de tonos. La división se controla por medio de un codificador digital que programa el contador para realizar las divisiones sucesivas según una secuencia específica a fin de generar las frecuencias correspondientes al tono deseado. Cada vez que se pulsa una tecla, se cierran dos contactos que ponen en funcionamiento el circuito generador de tonos multifrecuencia. Un codificador detecta los cierres de contactos y establece adecuadamente las divisiones sucesivas. El circuito incluye también un convertidor digital-analógico, que traslada las señales digitales correspondientes a frecuencias obtenidas por divisiones sucesivas, a señales alternas que son los tonos que se escuchan al marcar.

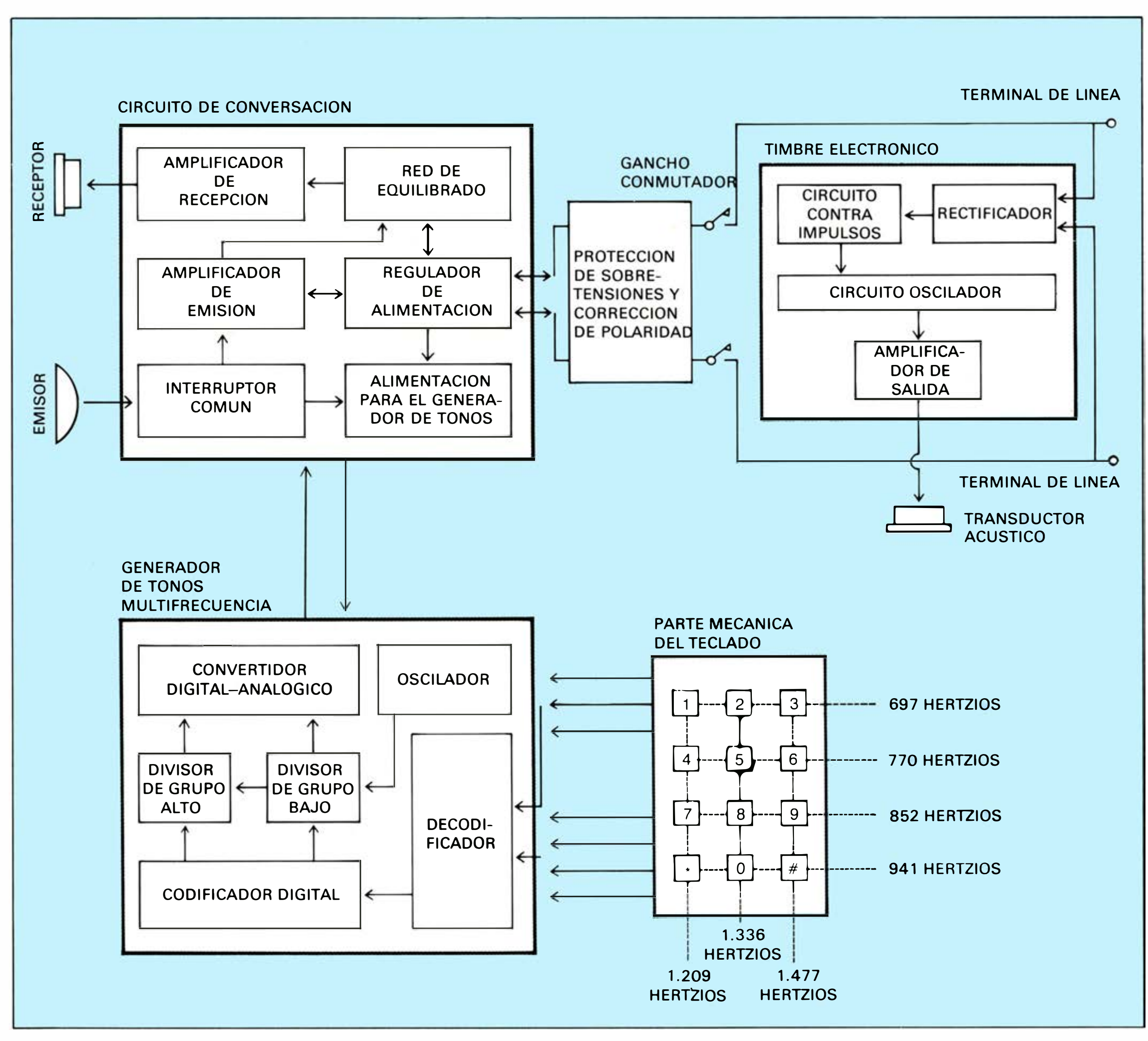
Sigamos una secuencia típica de marcación, para ilustrar la interacción del circuito generador de tonos multifrecuencia con el de conversación. Cuando se pulsa una tecla, el decodificador del generador de tonos envía una "señal común de interrupción" hacia el circuito de

conversación. Esta señal común de interrupción, denominada así porque realiza un número idéntico de funciones cualquiera que haya sido la tecla pulsada, permite que de nuevo se alimente el generador, iniciándose así la oscilación del reloj y el funcionamiento de los circuitos divisores. Simultáneamente, la señal común de interrupción corta el circuito de micrófono, y envía las señales multifrecuencia al amplificador-emisor. El circuito de micrófono se interrumpe porque el envío de señales vocales por línea, conjuntamente con tonos multifrecuencia, podría alterar el funcionamiento correcto del receptor de tonos de la central de conmutación. La señal común de interrupción reduce también adecuadamente la ganancia del receptor, para que el usuario no perciba un fuerte impacto sonoro a través del auricular. Cuando se libera la tecla, la secuencia completa se realiza en sentido inverso.

El circuito de timbre, no actúa en realidad como un timbre. En su lugar genera dos tonos cuyas frecuencias se encuentran bastante próximas. (La selección de las frecuencias presenta un problema de reacciones humanas. Por un lado, el sonido debería ser subjetivamente agradable, pero también lo suficientemente irritante como para estimular al usuario a contestar la llamada, en lugar de demorarse por disfrutar del sonido.)

Las funciones básicas de un teléfono, conversación, señalización y timbre fueron todas ellas desarrolladas por Alexander Graham Bell y sus contemporáneos antes del año 1890. Después de estos diseños básicos, la primera innovación significativa en el teléfono ocurrió unos 25 años más tarde con la aparición del nuevo método de señalización, el disco. El cambio se produjo no tanto por razones tecnológicas (realmente los sistemas manuales eran mejores que los sistemas de marcación por disco), sino más bien debido a que la rápida expansión de los sistemas telefónicos era acompañada por una progresiva escasez de operadoras.

La aparición del teléfono con micrófono y auricular montados en una sola pieza (microteléfono) tuvo lugar más tarde. Aunque concebida antes de 1900, no se introdujo a gran escala hasta finales de 1920. ¿Por qué se demoró tanto la introducción de una mejora tan obvia? El problema era muy conocido por los niños de la época, que con un aparato "dos piezas" podían hacerlo "cantar" al



Las funciones del teléfono electrónico se realizan en tres circuitos integrados, a saber: circuito de conversación, generador de tonos multifrecuencia y timbre electrónico

aproximar el auricular al micrófono. El pitido se producía por acoplamiento acústico entre los transductores a través del circuito de tono local. El efecto se intensificó cuando se hicieron los primeros intentos de colocar el auricular y micrófono en la misma pieza, porque así el sonido podía acoplarse directamente a través del mango del microteléfono. La producción de un microteléfono poco ruidoso absorbió una considerable labor de investigación.

Posteriores desarrollos, principalmente en la producción de cápsulas emisoras y receptoras, condujeron a la introducción del modelo 500, que es el habitual en Estados Unidos. En el año 1963 apareció el aparato de teclado, que consta esencialmente de los mismos componentes, a excepción del disco, que el modelo 500. No obstante, el aparato de teclado representa el comienzo de la era de los teléfonos electrónicos, ya que incorpora un transistor en el circuito generador de tonos multifrecuencia. Un diseño noruego, aparecido en 1967, representó un significativo despegue con respecto al aparato convencional, al sustituir el micrófono de carbón por un transductor electrodinámico más un circuito electrónico de amplificación. Un perfeccionamiento posterior fue el circuito de conversación electrónico, incorporado a un teléfono danés en el año 1973.

Veamos las ventajas de los teléfonos electrónicos. El micrófono de carbón de los teléfonos convencionales presenta varios inconvenientes. Su sensiblidad varía considerablemente en función de la corriente de bucle, de su período de vida y del trato que haya recibido. (Algunas personas todavía golpean el microteléfono con la esperanza de que sacudiendo la granalla de carbón se reduzca la distorsión de la voz y aumente el nivel de salida del micrófono.) La sustitución del emisor de carbón por uno lineal de tipo electrodinámico resuelve estos inconvenientes y dota al teléfono de una calidad de voz equiparable a la alta fidelidad.

El teléfono electrónico realiza la ecualización o equilibrado mucho mejor que el aparato 500. La razón se debe, en parte, a la eliminación de la pérdida de sensibilidad del micrófono. No obstante el teléfono electrónico compensa casi totalmente la atenuación de bucle largo. Por último el teléfono electrónico se comporta mejor en la eliminación del eco en bucles cortos.

Otras ventajas de los teléfonos electrónicos todavía no se han explotado. En un bucle largo el aparato electrónico recibe del orden de cinco veces más potencia que la requerida y 100 veces más

en un bucle corto. Si los sistemas de alimentación pudieran rediseñarse adecuadamente, en las centrales de conmutación se obtendría un ahorro de equipo instalado y de potencia consumida.

Los beneficios económicos inmediatos que puede ofrecer un teléfono electrónico surgen de la reducción sustancial del tamaño y número de componentes individuales requeridos. El ejemplo más evidente lo constituye el teclado. La unidad de teclado utilizada en el teléfono convencional consta de más de 120 elementos. La razón se debe a que la función común de interrupción se realiza mecánicamente utilizando un complicado sistema de palancas y muelles. También el circuito eléctrico necesita dos bobinas con núcleo de ferrita que, junto con un transistor, constituyen el circuito oscilador, resultando así un conjunto voluminoso y caro. Además las bobinas tienen que ser ajustadas individualmente durante su fabricación, a fin de centrar las diferentes frecuencias en los estrechos márgenes de tolerancia especificados. La función común de interrupción se realiza electrónicamente en el circuito de conversación. En consecuencia, la unidad de teclado es mucho más simple, tanto mecánica como eléctricamente, habiéndose reducido a muy pocos elementos.

El timbre electrónico, el circuito de conversación y los circuitos de protección tienen también menos elementos, y más reducidos, que sus homónimos en el aparato convencional. El ahorro total conseguido en el teléfono electrónico es del orden de 180 elementos. Teniendo en cuenta que la mayoría de los elementos son condensadores, resistencias y circuitos integrados, esto es, componentes de bajo coste, pueden aplicarse en la fabricación de aparatos telefónicos los métodos desarrollados por la industria electrónica para el ensamblaje automático.

Aunque el teléfono electrónico se debe principalmente a los reciente avances experimentados en la tecnología de circuitos integrados y en las características de los semiconductores, también las innovaciones habidas en otros componentes desempeñan un notable papel en la reducción de costes. Por ejemplo, la frecuencia del oscilador maestro empleado en el generador de tonos multifrecuencia debe tener una estabilidad de 0,2 por ciento. Antes, este estricto requerimiento solamente podía lograrse por medio de un resonador de cristal de cuarzo cortado con precisión. Recientemente se han desarrollado resonadores de cerámica mucho más baratos. El coste de otro componente, el que protege el teléfono contra sobretensiones, se ha reducido a la tercera parte en los dos últimos años.

Se puede añadir un visualizador similar al de una calculadora electrónica. Con dicho dispositivo se puede contabilizar el tiempo empleado en llamadas de larga distancia, verificar los números almacenados y realizar operaciones aritméticas con el propio teclado. Otra opción para el usuario de un teléfono electrónico es la de poder variar el volumen y sonido del timbre para adaptarlo a cualquier situación. Consiguiendo, por ejemplo, un tono suave y melodioso para el cuarto de baño y otro fuerte y estridente para un taller. Una característica que puede incorporarse fácilmente en el módulo electrónico de un teléfono residencial es la "retención", facilitando al usuario la posibilidad de transferir una llamada desde una extensión a otra, sin necesidad de volver posteriormente a la primera extensión para colgar el microteléfono. Finalmente, podría incorporarse en todos los teléfonos, la facilidad "manos-libres" que permitiría al usuario hablar y escuchar sin utilizar microteléfono. (P. P. Luff)

Imanes microbianos

Acaba de registrarse un nuevo avance en el estudio de las bacterias que se orientan hacia el norte, por dos científicos dedicados a su investigación. La historia se remonta al otoño de 1975, cuando Richard P. Blakemore, que trabaja hoy como microbiólogo en la Universidad de New Hampshire, dio cuenta de un curioso descubrimiento. A medida que examinaba lodos recogidos en lagunas marinas y de agua dulce en las proximidades de Woods Hole, Massachusetts, Blakemore observó en su microscopio que ciertas clases de bacterias nadaban uniformemente hacia el norte al ser separadas de sus sedimentos. Descubrió que la dirección en que nadaban podía cambiarse adecuadamente moviendo un pequeño imán por encima de la placa del microscopio.

Blakemore se unió con Adrianus J. Kalmijn, especialista en biofísica sensorial de los vertebrados marinos, del Instituto Oceanográfico de Woods Hole, para probar que estas bacterias también se mueven hacia el norte en un campo magnético generado artificialmente, de una intensidad comparable a la del campo magnético terrestre. Además, Blakemore y Kalmijn pudieron comprobar que cuando las bacterias están fuera del entorno del laboratorio (en los bosques cercanos, libres de las distorsiones del campo geomagnético causadas por el hombre), siguen realmente las líneas de fuerza del campo geomagnético, muy inclinadas. La tendencia hacia el norte por parte de las bacterias era particularmente clara cuando se invertía el campo magnético del ambiente mediante unas grandes bobinas Helmholtz: los organismos unicelulares cambiaban inmediatamente su rumbo, haciendo un giro de 180 grados que tenía una anchura de varias veces su diámetro celular. Se volvían a alinear con el campo en cuestión de segundos, esta vez orientados en el sentido opuesto. Ambos investigadores supusieron que estas bacterias tenían un momento magnético dipolar permanente, basándose para ello en su comportamiento magnético y en el descubrimiento de Blakemore de que contenían partículas ricas en hierro; en otras palabras, parecían imanes en miniatura.

El último esfuerzo acometido por Kalmijn y Blakemore consistió en demostrar experimentalmente esta nueva hipótesis sometiendo la bacteria a un impulso magnético relativamente fuerte. Los impulsos fueron superpuestos al campo magnético estático terrestre, y la componente vertical de dicho campo fue anulada a fin de que la bacteria sólo avanzara en el plano horizontal. Cuando el impulso era de sentido contrario a la componente horizontal del campo terrestre (y era suficientemente intenso), los investigadores observaron que el campo aplicado motivaba al instante que la bacteria girase y nadase en la dirección opuesta (es decir, recto hacia el sur), fenómeno que podría esperarse a partir de la inversión de sus momentos dipolares magnéticos intrínsecos.

Se vio que era necesario un impulso de unos 375 o 400 gauss de intensidad, aproximadamente, para afectar al 50 por ciento de las muestras de agua dulce. Para las muestras marinas, el correspondiente valor era significativamente más alto: entre 525 y 550 gauss. Según la intensidad del impulso magnético aplicado, las bacterias permanecían orientadas hacia el norte o cambiaban y se orientaban hacia el sur: no había ninguna que quedara totalmente sin polarizar. A este respecto, Kalmijn y Blakemore señalan que los organismos analizados presentan las propiedades características de los dominios magnéticos elementales de un material ferromagnético. Concluyen que, aunque esté por determinar la función biológica de esta respuesta, el comportamiento magnético de las bacterias en cuestión constituyen una prueba directa de la orientación ferromagnética en la naturaleza.



El flujo de energía en un ecosistema de bosque

La vida en la tierra se sustenta por una cantidad finita de energía solar que fijan las plantas verdes. El estudio cuantitativo del flujo de energía de un bosque ha mostrado de qué manera se distribuye y controla dicha energía

James R. Gosz, Richard T. Holmes, Gene E. Likens y F. Herbert Bormann

La ecología, como la economía, trata del movimiento de productos valiosos a través de una compleja red de productores y consumidores. De la misma manera que una economía funciona a base de dinero, un ecosistema funciona a base de energía, toda ella procedente en principio del sol. Un ecosistema, que es una unidad funcional básica de la naturaleza, comprende un grupo de organismos vivos y el ambiente físico y químico en el que éstos viven. Para nuestros propósitos, puede considerarse formado por plantas, animales, detritos orgánicos, nutrientes disponibles, minerales del suelo, agua y gases, todo ello ligado por redes tróficas y flujos de energía y de nutrientes. Los organismos productores y los organismos consumidores interactúan en forma tal que se autorregulan, por lo general en relación con la cantidad total de energía disponible en el ecosistema.

Aunque puede parecer que el mundo natural recibe un flujo de capital energético virtualmente ilimitado en forma de radiación solar, su presupuesto de energía es en realidad bastante reducido. Esta paradoja puede explicarse por el hecho de que los organismos vivos no utilizan la energía solar directamente; sólo tienen acceso a la pequeña fracción de energía solar que es convertida por las plantas verdes en una forma de energía química almacenada, mediante el proceso de la fotosíntesis. La materia orgánica fijada por las plantas es utilizada por los consumidores animales; las plantas son también consumidores en la medida en que utilizan para su propio mantenimiento parte de la energía almacenada en sus tejidos. Los animales están asociados a una red trófica de herbívoros (en la que la fuente de energía es tejido vegetal vivo) o a una red trófica de detritívoros (en la que la fuente está constituida por tejidos muertos).

Puesto que la mayoría de los consumidores no aprovechan eficazmente dicha energía química, la cantidad finita de energía solar fijada en un ecosistema limita severamente el número de plantas y de animales que éste puede sostener. En consecuencia, estos organismos han tenido que desarrollar distintas estrategias para obtener su parte de la energía disponible. El análisis que nuestro grupo ha efectuado del presupuesto de energía de un ecosistema de bosque, al nordeste de los Estados Unidos, ha empezado a revelar con cierto detalle de qué manera los organismos distribuyen y controlan el movimiento de energía a través del sistema.

Hay tres maneras generales de investigar la dinámica de la energía en la naturaleza, cada una de ellas con sus puntos débiles y con sus ventajas. El enfoque más específico es el análisis de la población, que estima el presupuesto de energía de una población dada en la naturaleza. El análisis se realiza determinando la cantidad de energía en forma de materia orgánica que se consume, asimila y excreta por los distintos organismos, más la cantidad de energía requerida para el crecimiento y las actividades de mantenimiento, como la respiración (la oxidación del alimento). Estos datos, que se obtienen en el laboratorio, se combinan después con estudios demográficos de la población en su hábitat natural al objeto de calcular la cantidad de energía que utiliza la población. Este enfoque, sin embargo, requiere mucho tiempo, y, por consiguiente, sólo se ha estudiado un número limitado de especies.

El segundo enfoque se centra en el análisis de las cadenas tróficas. En él se investiga la energética de la población de una o dos especies a distintos niveles de una cadena trófica. Por ejemplo, una especie vegetal puede servir de alimento para una especie de insecto, que a su vez nutre a una especie de ave. Cada especie se considera como si formara el único eslabón de la cadena trófica, y se ignoran las rutas alternativas del flujo de energía. Este enfoque proporciona información detallada acerca de la manera en la que las especies interactúan para distribuirse los recursos de energía disponibles, pero revela muy poco sobre la distribución de energía en el interior de las complejas redes tróficas de un ecosistema.

El tercer enfoque, que es el más amplio, incide en el análisis del ecosistema. Esta aproximación cuantifica la cantidad de energía transferida entre los compartimientos consumidores de un ecosistema. Puesto que pueden calcularse las entradas y salidas de cada compartimiento, cabe esbozar un balance del flujo energético. A fin de dotar a este balance de una estructura contable, se definen fronteras topográficas para el ecosistema que se está investigando, de modo que el flujo de energía pueda expresarse por unidad de tierra o de agua y por unidad de tiempo. La energía suele expresarse en kilocalorías por metro cua-

ABSORCION DE LA RADIACION SOLAR en un bosque caducifolio a latitudes de la zona templada, en la página opuesta. La fotografía aérea se realizó con una emulsión especial sensible a las longitudes de onda del infrarrojo cercano. El follaje de las plantas absorbe la luz, ante todo, a las longitudes de onda azul y roja, y refleja el resto; así los árboles sanos aparecen rojos en esta fotografía porque reflejan el infrarrojo cercano. Los árboles que aparecen verdes han sido parcialmente desnudados de sus hojas por orugas de polilla lagarta, reduciendo así su capacidad para absorber y fijar la radiación solar. En los años en los que los insectos defoliadores son especialmente abundantes, pueden constituir un impacto importante en el funcionamiento del ecosistema de bosque al afectar la tasa y la cantidad de fijación de energía. La fotografía fue cedida por Robert L. Talerico.

drado y por año. Una kilocaloría equivale a 1000 calorías, siendo una caloría la cantidad de energía necesaria para aumentar en un grado Celsius la temperatura de un gramo de agua.

El ecosistema que hemos estudiado se halla en el bosque experimental del Hubbard Brook, encomendado al U.S. Forest Service en el White Mountain National Forest, en New Hampshire. En el interior del bosque experimental hay varias cuencas hidrográficas asentadas sobre un lecho de roca impermeable y divididas por cerros (elevaciones topográficas). Una de ellas, designada como Cuenca 6, se mantiene en condiciones naturales, de manera que puede servir como base de referencia para estudios experimentales de hidrología y de ciclos biogeoquímicos. El análisis del flujo de energía se realizó en esta cuenca, que tiene una superficie total de 132.300 metros cuadrados. La vegetación de la cuenca natural es un bosque renacido, bien desarrollado, de arce de azúcar *(Acer saccharum),* haya americana *(Fagus grandifolia)* y abedul amarillo *(Betula alleghaniensis),* y no ha sido dañado por fuegos o talas desde 1919, año en el que la industria maderera extrajo gran parte de la vegetación original.

Aunque la Cuenca 6 tiene fronteras topográficas definidas, que corresponden al lecho rocoso subyacente y a las divisorias montañosas, es "abierta" en el sentido de que la supervivencia depende de la entrada y salida continua de energía, en forma de radiación y de materia orgánica. La materia orgánica que contiene energía puede ser transportada a través de las fronteras del ecosistema por medio de fuerzas meteorológicas, como la precipitación y el viento, por medio de fuerzas geológicas, como el agua corriente, y mediante vectores biológicos, como el desplazamiento de los animales. Todas estas rutas de transporte deben tenerse en cuenta al describir cuantitativamente el flujo de energía que atraviesa el ecosistema.

Medir el movimiento de energía a través de las fronteras del ecosistema suele ser una tarea difícil, pero en el ecosistema de nuestra cuenca varios factores contribuyeron a su simplificación. En primer lugar, puesto que el ecosistema se asienta sobre un lecho rocoso impenetrable, la salida geológica de materia orgánica podía cuantificarse con facilidad porque ésta aparecía inevitablemente en los arroyos que drenan la cuenca. Además, el transporte biológico a través de las fronteras del ecosistema se consideró despreciable porque los animales consumían y descargaban materia orgánica al azar al moverse hacia dentro del ecosistema desde bosques similares en las regiones vecinas o hacia fuera del mismo. Por todo ello, nos bastó con medir la entrada de energía solar, la entrada meteorológica y la salida geológica para estimar el flujo de energía total que atravesaba el ecosistema.

La cantidad de radiación solar recibida por un ecosistema a lo largo de un día determinado depende de la latitud, el clima y la época del año. Asimismo, puede haber variaciones de un año para otro como consecuencia de los cambios registrados en las condiciones climatológicas. Basamos nuestro análisis del flujo de energía en datos de radiación solar obtenidos desde el primero de junio de 1969 hasta el 31 de mayo de 1970, un año típico para la región del Hubbard Brook. Durante ese período, alcanzaron la Cuenca 6 un total de 1.254.000 kilocalorías por metro cuadrado de radiación solar, de las que 10.400 kilocalorías por metro cuadrado (el 0,8 por ciento) fueron fijadas en forma química por las plantas, mediante la fotosíntesis. Las plantas de este bosque necesitan el 55 por ciento de la energía que fijan para su propio sostenimiento, de manera que la cantidad real de nuevo tejido vegetal que se formó ese año (la producción primaria neta) fue de unas 4680 kilocalorías por metro cuadrado, es decir, alrededor del 0,4 por ciento de la entrada anual total de energía solar. Este porcentaje es la eficiencia de la producción primaria neta.

El cálculo de la eficiencia de la producción primaria neta es inexacto por dos razones. En primer lugar, induce a confusión expresar la eficiencia de la producción de los vegetales en función de la radiación solar anual total porque las plantas de esta región carecen de hojas o se hallan en estado de vida latente a lo largo de ocho meses durante el año. Aunque, en algunas especies, durante los meses de invierno existe ya un nivel bajo de actividad fotosintética, de hecho todo el crecimiento tiene lugar en la primavera tardía y el verano. Por ello, se prefiere calcular la eficiencia de la fotosíntesis sobre la base de que el crecimiento de las plantas se halla limitado al período que va de junio a septiembre. La radiación solar total durante esos cuatro meses fue de 480.000 kilocalorías por metro cuadrado, lo que significa que la producción primaria neta de 4680 kilocalorías por metro cuadrado correspondería a cerca del 1 por ciento de la entrada de energía solar, valor que es algo superior a nuestra estima inicial del 0,4 por ciento.

Ocurre también que, aunque el valor energético correspondiente a la radia-

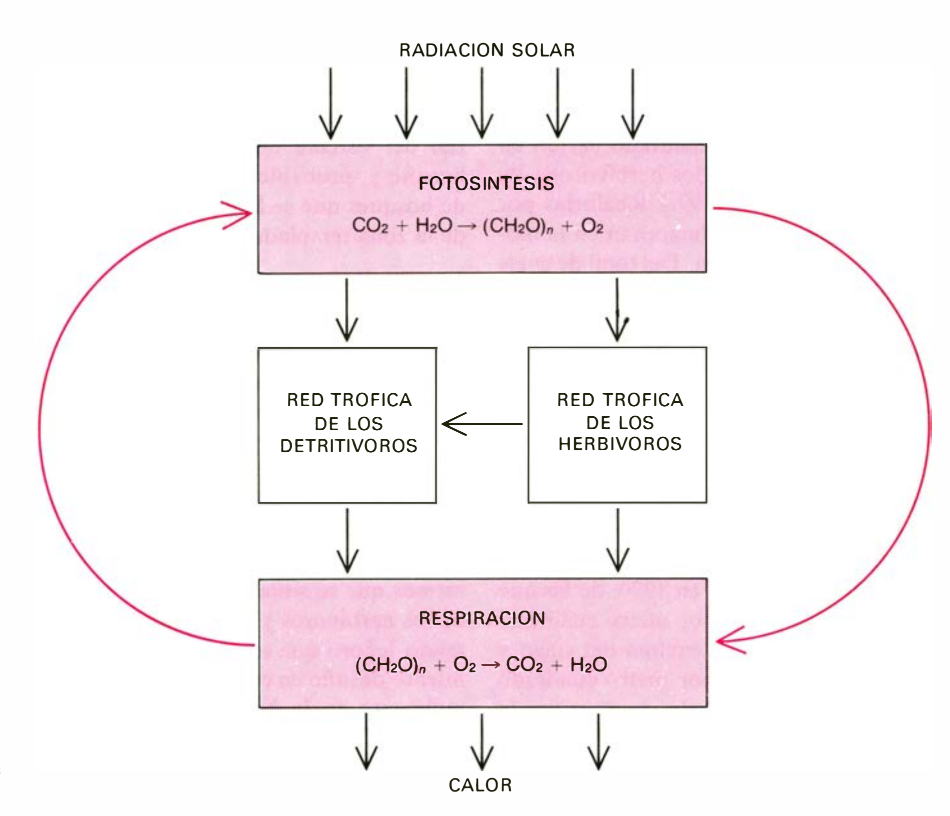
ción solar total comprende las energías de todas las longitudes de onda de radiación electromagnética que, desde el sol, llegan a la tierra, el proceso fotosintético no utiliza con la misma eficiencia todas las longitudes de onda. La clorofila absorbe fuertemente la luz de las regiones azul y roja del espectro visible, y refleja la mayor parte de luz verde, mientras que los otros pigmentos de las hojas de las plantas, como los carotenos y las xantofilas, absorben las longitudes de onda del verde y transfieren parte de la energía absorbida a las moléculas de clorofila, que pueden utilizarla en la fotosíntesis. El efecto combinado de los distintos pigmentos determina que la mitad de la energía presente en las longitudes de onda de la radiación solar, aproximadamente, se halle disponible para su conversión en forma química. Teniendo todo esto en cuenta en nuestros cálculos, resulta que la fotosíntesis es ligeramente más eficiente de lo que se ha descrito más arriba: durante la estación de crecimiento, alrededor del 2 por ciento de la energía solar de las longitudes de onda apropiadas se convierte en materia orgánica.

El hecho de que sólo una fracción minúscula de la radiación solar que alcanza el ecosistema se convierta en nuevo tejido vegetal no significa que la comunidad vegetal sea derrochadora o que el resto de la energía no se aproveche. Gran parte de la radiación solar genera calor cuando es absorbida por el suelo u otros componentes del ecosistema, y sin calor que mantuviera los niveles normales de metabolismo no podría actuar la fotosíntesis.

Otro proceso dependiente de la energía, y crucial para el funcionamiento de las plantas, es la transpiración: ésta es el movimiento de agua desde el suelo a través de la planta y su evaporación en la superficie de la hoja. Se ha demostrado que la transpiración es un factor importante para el crecimiento normal de la planta. Elimina agua sin erosionar el paisaje o arrastrar consigo nutrientes esenciales. La cantidad total de agua eliminada conjuntamente del ecosistema por la evaporación (no debida a los organismos) y la transpiración (debida a ellos) es la evapotranspiración, y puede calcularse en nuestra cuenca restando la salida de agua del arroyo de la entrada de agua por precipitación.

Según nuestros cálculos, cada año la evapotranspiración elimina un volumen de agua equivalente a una capa de 49,4 centímetros de altura a lo largo de toda la superficie de la cuenca. A la temperatura normal de nuestro bosque, se necesitan unas 585 calorías aproximadamente para evaporar un gramo de agua, de modo que la evapotranspiración requiere 288.990 kilocalorías por metro cuadrado y por año, es decir, alrededor del 23 por ciento de la entrada anual total de energía solar. En la estación de crecimiento, la evapotranspiración elimina 34,2 centímetros de agua y consume 200.000 kilocalorías por metro cuadrado, o sea cerca del 42 por ciento de la radiación solar del período de cuatro meses. Así, con la excepción de las longitudes de onda verde e infrarroja, que son reflejadas por las plantas, la mayor parte de la radiación solar incidente es responsable de la creación de las condiciones ambientales que permiten que la comunidad vegetal fije en materia orgánica 4680 kilocalorías de energía por metro cuadrado por año.

La otra entrada principal de energía en el ecosistema es el transporte meteorológico de materia orgánica, ya sea en forma particulada a través del viento, ya en forma química disuelta en la lluvia y la nieve. La precipitación que cae en las cuencas del Hubbard Brook contiene una media de unos tres miligramos de carbón orgánico por litro, lo que corresponde a una entrada de energía de unas 35 kilocalorías por metro cuadrado y por año. Aunque la entrada meteorológica es pequeña si se la compara con la entrada de radiación solar, representa más de cinco veces la energía contenida en la materia orgánica que abandona el ecosistema en el agua del arroyo. Además, puesto que esta energía está en forma química, se halla disponible directamente para uso de los consumidores. Una vez la energía se ha convertido en forma química por la fotosíntesis, o ha sido transportada hasta el ecosistema por el viento o la precipitación, puede entrar en la red trófica de los herbívoros a través del consumo directo de tejido vegetal vivo o bien entrar en la red trófica de los detritívoros del suelo. Alternativamente, puede abandonar el ecosistema a través del movimiento animal, como la migración de las aves, en forma de materia orgánica disuelta en el agua del arroyo, como materia orgánica volátil en el aire o bien en forma de calor. Si la producción vegetal es mayor que la cantidad de materia orgánica que entra en las distintas redes tróficas y rutas de exportación, el exceso se acumulará en la biomasa de las plantas vivas, entendiendo por ésta la cantidad total de tejido vegetal vivo en el ecosistema.



LAS RUTAS BASICAS del flujo de energía a través de un ecosistema se originan en su totalidad en la radiación solar; de ésta, las plantas verdes convierten un pequeño porcentaje en materia orgánica a través del proceso de la fotosíntesis. Esta materia orgánica puede consumirse en forma de tejido vegetal por los animales de la cadena trófica de los herbívoros, o bien en forma de tejido muerto por los descomponedores de la cadena trófica de los detritívoros; estos animales proporcionan a su vez el alimento de base para diversos carnívoros. La materia orgánica asimilada, que no es utilizada para el crecimiento de los organismos individuales o de sus poblaciones, es "quemada" por el proceso de respiración, que mantiene el metabolismo de plantas y animales. Al perderse la energía respirada en forma de calor, todo ecosistema necesita un flujo continuo de energía solar.

Hemos calculado que, de la producción primaria anual total de 4680 kilocalorías por metro cuadrado, 3481 kilocalorías por metro cuadrado entran en las redes tróficas de los herbívoros y de los detritívoros y 1199 kilocalorías por metro cuadrado se almacenan en la biomasa de la vegetación. Del total de energía almacenada, 952 kilocalorías por metro cuadrado se hallan por encima del suelo y 247 kilocalorías por metro cuadrado por debajo del mismo. De ahí resulta claro que este bosque renacido se halla todavía creciendo y aumentando su biomasa después de las operaciones madereras realizadas a principios de siglo.

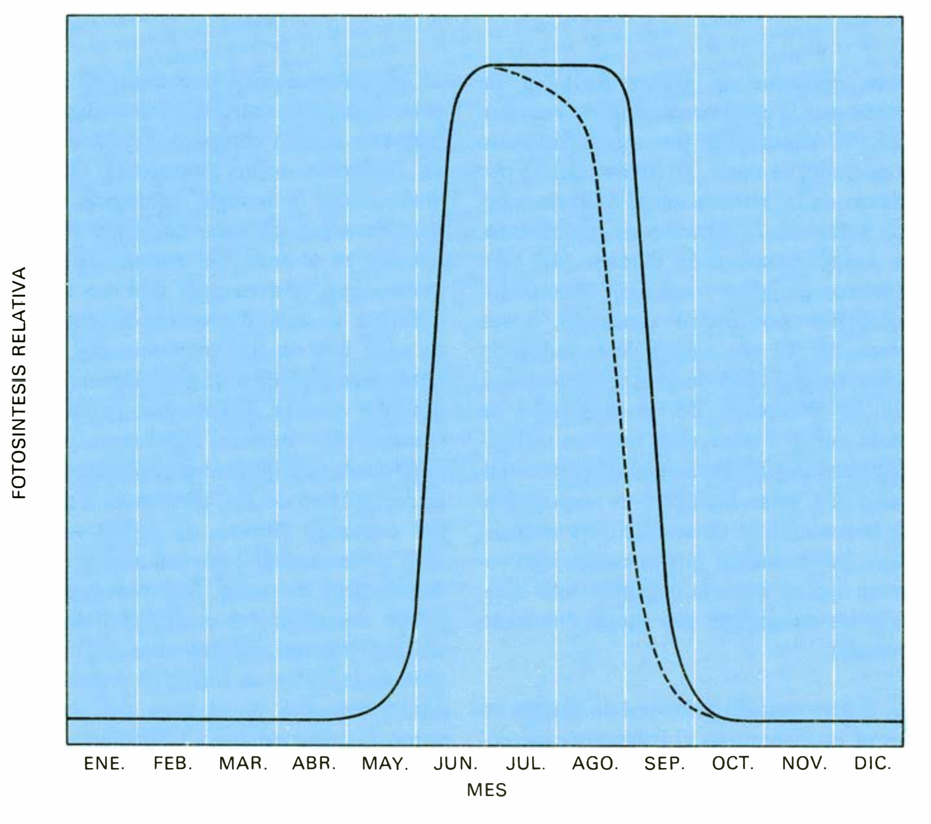
La biomasa total de vegetación viva del bosque contenía 71.420 kilocalorías por metro cuadrado en 1970, de las que 59.696 kilocalorías por metro cuadrado se encontraban por encima del suelo y 11.724 kilocalorías por metro cuadrado se hallaban bajo el suelo. Aunque puede parecer que la biomasa viva del bosque es la forma de almacenamiento más importante de materia orgánica y de energía, no es así; la materia orgánica que se encuentra sobre el suelo del bosque y mezclada con el suelo contiene 1,7 veces esta energía. El suelo del bosque tenía 34.322 kilocalorías por metro cuadrado en 1970; la materia orgánica en el suelo hasta una profundidad de 36 centímetros contenía 88.120 kilocalorías por metro cuadrado. Estas cifras indican la magnitud del sistema de los detritos en este bosque y, probablemente, en la mayoría de bosques que se hallan en las latitudes de la zona templada.

El tejido vegetal vivo que está a disposición de los consumidores en el ecosistema de bosque es de tres tipos: semillas y frutos, hojas y tejido leñoso. La producción de semillas (80 kilocalorías por metro cuadrado y por año) y la producción de follaje (1485 kilocalorías por metro cuadrado por año) son fracciones de la producción primaria neta anual. Estos tejidos son consumidos por organismos que se sitúan en las redes tróficas de los herbívoros y de los detritívoros. El tejido leñoso que se produce en el crecimiento del año en curso se almacena normalmente en la biomasa de la vegetación; la fracción de tejido leñoso que es utilizada por los consumidores consiste primariamente en tejido muerto de tallos, ramas y raíces, que representa la producción vegetal de años anteriores.

Los principales consumidores en la red trófica de los herbívoros son insectos folífagos, o comedores de hojas, sobre todo orugas. Aunque ratones, ardillas de tierra (chipmunks), ciervos y liebres variables consumen parte del tejido foliar que se halla en el suelo o cerca de él, la amplia variación en las cantidades consumidas en los distintos años se relaciona principalmente con las fluctuaciones de la población de orugas. De las 1485 kilocalorías por metro cuadrado y por año que representa la producción neta de follaje en el bosque del Hubbard Brook, estos animales consumen cantidades que van desde menos del 1 por ciento en la mayoría de años hasta más del 40 por ciento en otros. Por ejemplo, desde 1969 a 1971, época en la que una especie de oruga folífaga (*Heterocampa guttivita* Walker) fue extraordinariamente abundante, en Hubbard Brook se consumió alrededor del 44 por ciento del tejido foliar total. En ciertas áreas de las cuencas y en otras regiones de los Estados Unidos nororientales, árboles y partes enteras de bosque fueron totalmente desnudados de sus hojas. Aunque estas graves pululaciones suelen ocurrir de manera ocasional en los bosques septentrionales, representan momentos en los que los insectos folífagos pueden tener una incidencia importante en el funcionamiento del ecosistema al afectar la tasa y la cantidad de fijación de energía.

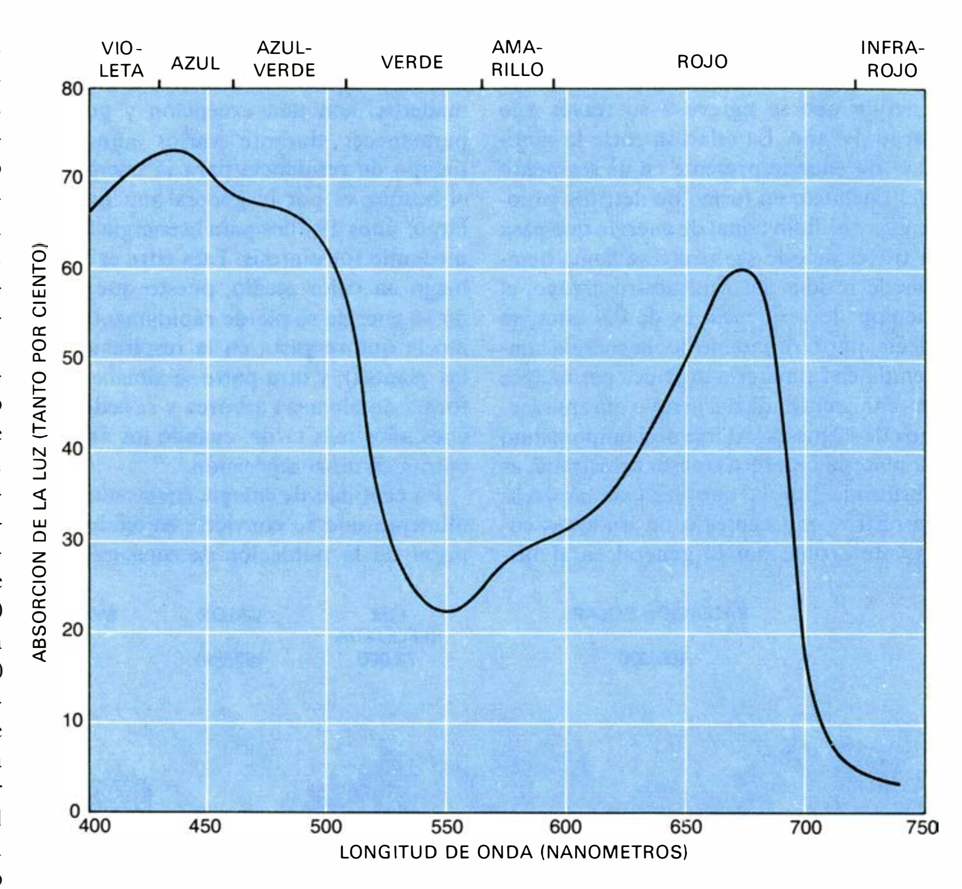
Año tras año, con pocas excepciones, cerca del 75 por ciento de la producción anual neta en nuestro ecosistema no es consumida por los animales de la red trófica de los herbívoros ni se acumula en la biomasa vegetal viva. Y así ocurre que cae al suelo del bosque y entra en la ruta energética de los detritos. La cantidad anual media de energía que entra en esta ruta se estima en 3505 kilocalorías por metro cuadrado. La caída de hojas, ramas y troncos de la bóveda arbórea es la vía principal de transferencia de energía, y representa un 83 por ciento del total. El 17 por ciento restante procede de diversas fuentes: muerte de las raíces (12 por ciento), caída de hojarasca de la vegetación baja (2 por ciento), materia orgánica arrastrada por la precipitación o lavada del follaje (2 por ciento), materiales fecales de los animales de la red trófica de los herbívoros (0,9 por ciento) y materia orgánica exudada por las raíces (0,1 por ciento). La muerte de los animales pertenecientes a la cadena trófica de los herbívoros añade solamente una kilocaloría por metro cuadrado y por año (0,03 por ciento).

La materia orgánica que cae al suelo del bosque es utilizada por los consumidores de la red trófica de los detritívoros; en ésta se encuentran las bacterias, los hongos y una gran variedad de invertebrados, como protozoos, milpiés y ciertas larvas de insectos. A su vez, estos



LA PRODUCCION FOTOSINTETICA de las plantas no es uniforme a lo largo de todo el año, sino que se concentra en el período de cuatro meses que va de junio a septiembre; durante los otros ocho meses, la mayoría de árboles de los bosques septentrionales se hallan en un estado latente, sin hojas. La curva continua indica la fotosíntesis relativa en el ecosistema bajo condiciones estivales favorables; la curva de trazos corresponde a condiciones de estrés, como humedad reducida. La eficiencia con la que las plantas convierten energía solar en materia orgánica debería calcularse sobre la base de la cantidad de energía solar recibida por las plantas durante la estación de crecimiento.

consumidores sirven de presa a los invertebrados carnívoros (como ciempiés, escarabajos y arañas) y a determinados vertebrados (como musarañas, salamandras, roedores y algunas aves). Puesto que cada año el 75 por ciento de la energía de la producción primaria neta entra en la ruta de los detritos, una cuestión importante que se plantea es si los organismos de la red trófica de los detritívoros consumen toda la entrada anual de energía, o bien si existe una acumulación de materia orgánica en el suelo del bosque. Un análisis de la caída de hojarasca y de las características del suelo del bosque en varios bosques caducifolios septentrionales, en New Hampshire, sugiere que, después de las actividades de tala, hay un claro descenso de la materia orgánica durante 15 a 20 años: a partir de entonces tiene lugar una acumulación neta durante unos 150 años. Aunque nuestro ecosistema de bosque se ha mantenido inalterado durante unos 60 años, el suelo del mismo está acumulando unas 150 kilocalorías por metro cuadrado y por año. La suma del aumento de la biomasa vegetal viva (1199 kilocalorías por metro cuadrado y por año) y del contenido energético del suelo del bosque (150 kilocalorías por metro cuadrado y por año) representa la producción neta total de materia orgánica del ecosistema.



EL ESPECTRO TOTAL DE ABSORCION para los distintos pigmentos presentes en el follaje de las plantas indica que cerca de la mitad de las longitudes de onda que forman la radiación solar se usan en la fotosíntesis. El follaje refleja más intensamente la luz en la región verde e infrarroja.

Una de las principales salidas de energía del ecosistema es la pérdida de materia orgánica a través de los riachuelos que drenan la cuenca. Puesto que este factor puede cuantificarse, nos permite evaluar la dinámica energética de la fracción acuática del ecosistema. Hay tres rutas para la entrada de energía en el ecosistema. Una pasa por la producción primaria de las plantas acuáticas del propio arroyo, que utilizan la radiación solar directamente. Nuestro arroyo, sin embargo, es similar a la mayoría de riachuelos de cabecera en regiones densamente forestadas, en el sentido de que es muy umbrío y posee una vida vegetal muy reducida, formada principalmente por musgos. De ello resulta que la contribución energética de las plantas acuáticas es de sólo 10 kilocalorías por metro cuadrado de lecho del arroyo por año, alrededor del 0,2 por ciento de la entrada total de energía en el arroyo.

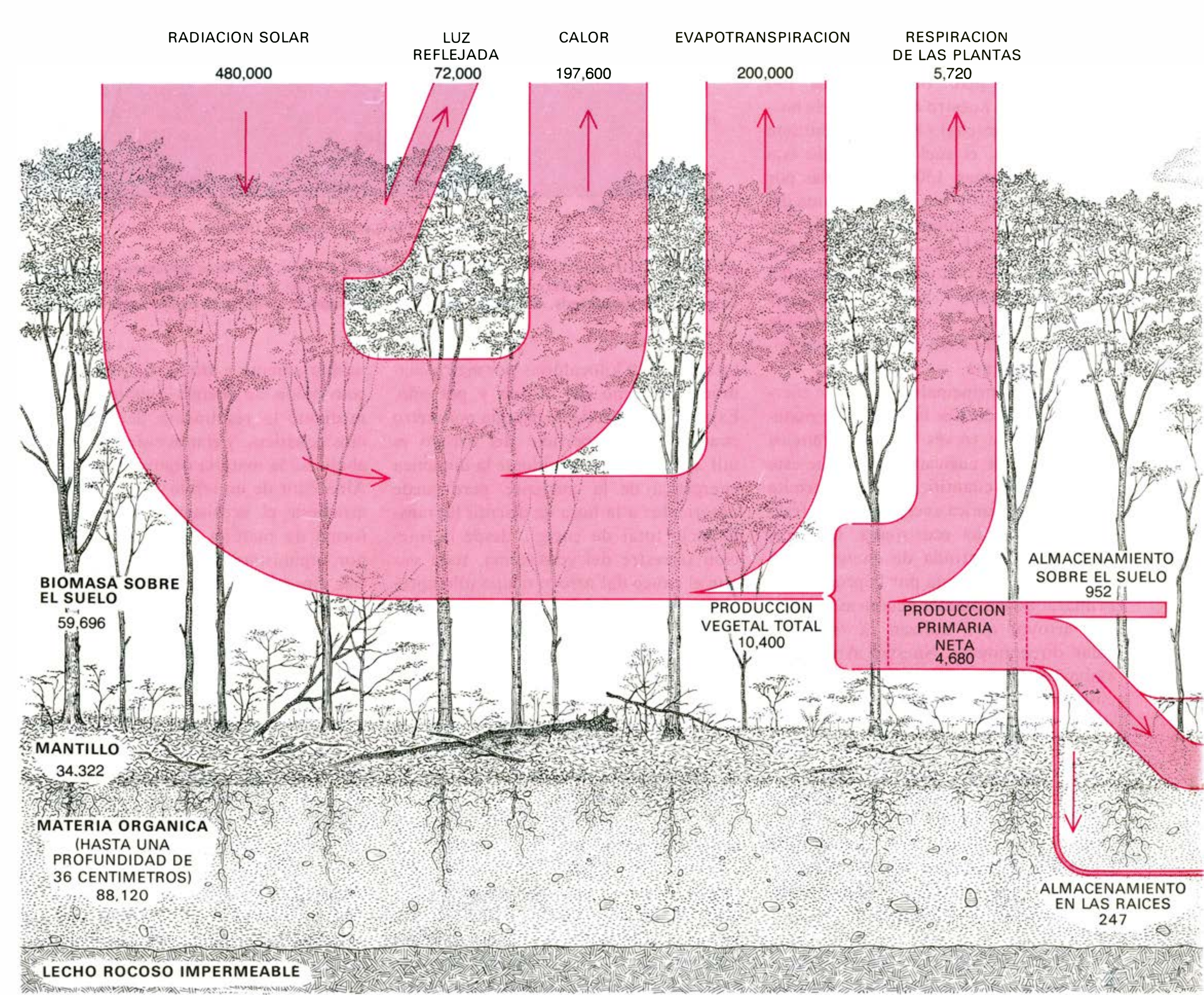
Las otras dos rutas (a saber: la transferencia de materia orgánica al arroyo desde el bosque y el suelo del bosque circundantes y el movimiento de materia orgánica particulada y disuelta hacia el cauce del riachuelo en el agua de drenaje) representan la mayor parte de la energía que entra en el arroyo, y suman un total de 6039 kilocalorías por metro cuadrado de lecho del arroyo y por año. Expresar la entrada de energía por metro cuadrado de superficie del arroyo es útil a la hora de considerar la dinámica energética de la corriente, pero puede desorientar a la hora de discutir la transferencia total de energía desde la fracción terrestre del ecosistema, toda vez que el cauce del arroyo ocupa sólo aproximadamente el 0,2 por ciento de la superficie de la cuenca. Aunque los aportes energéticos que implica la caída de hojarasca de la bóveda arbórea y de la precipitación son directos, en el sentido de que un metro cuadrado de superficie del arroyo equivale, desde el punto de vista de la funcionalidad, a un metro cuadrado de tierra, la hojarasca que cae al arroyo y el agua que se drena hacia el mismo se hallan sujetas a un efecto de concentración. Debido a la alta relación superficie de tierra/superficie de arroyo, pérdidas relativamente pequeñas de cada metro cuadrado de superficie terrestre representan contribuciones elevadas a la energía del arroyo. En el ecosistema del Hubbard Brook, la entrada total de energía a la corriente de agua desde el bosque oscila en torno al 1 por ciento de la productividad neta de la vegetación en el bosque.

La salida de energía del riachuelo fluye primariamente a través de dos rutas: la conversión de energía química en calor, mediante la respiración de los organismos acuáticos, y el movimiento corriente abajo de la materia orgánica en el agua. Alrededor de un tercio de la energía que atraviesa el ecosistema del arroyo en forma de materia orgánica es utilizada por organismos acuáticos, como bacterias, insectos y truchas. Puesto que la pérdida de energía que representa la respiración de estos organismos es mucho mayor que la energía fijada por la fotosíntesis de las plantas acuáticas, los consumidores del arroyo dependen de la entrada de energía de las áreas de tierra circundantes. La materia orgánica portadora de energía que es arrastrada por el agua en movimiento (4004 kilocalorías por metro cuadrado y por año) sirve aparentemente como fuente de energía para organismos acuáticos que viven a cierta distancia corriente abajo, donde las entradas energéticas a partir de las áreas terrestres adyacentes son proporcionalmente menores.

La mayor parte de la energía almacenada en el arroyo en un momento dado se mantiene en la materia orgánica del depósito de detritos (4730 kilocalorías por metro cuadrado de lecho del arroyo), y sólo hay 14,5 kilocalorías por metro cuadrado en la biomasa viva.

La energía almacenada en el arroyo es poca si se la compara con la cantidad de energía que se mueve a su través a lo largo del año. La relación entre la cantidad de energía presente en un segmento del riachuelo en forma de detritos y biomasa y el flujo anual de energía que pasa a través de este segmento se llama tiempo de residencia. En nuestro arroyo, el tiempo de residencia es de 0,8 años, es decir, por término medio, la energía contenida en la materia orgánica permanece en una zona dada del arroyo durante menos de 10 meses. Al irse descomponiendo la materia orgánica o al ser arrastrada, es sustituida por la que llega de las áreas terrestres adyacentes o de las áreas corriente arriba, por lo general en el mismo año. (Las presas de restos, cuyo armazón estructural lo constituyen grandes maderos, son una excepción y pueden permanecer durante varios años.) El tiempo de residencia para la energía en el bosque es por lo general mucho más largo: unos 19 años para la energía fijada mediante fotosíntesis. Esta cifra es desde luego un valor medio, puesto que parte de la energía se pierde rápidamente (como la que se gasta en la respiración de las plantas), y otra parte se almacena en forma de biomasa arbórea y se cede muchos años más tarde, cuando los árboles entran en descomposición.

La cantidad de energía encerrada en el alimento que se convierte en tejido corporal de la población de consumidores depende de varios factores. Después de la ingestión, cierta cantidad de energía no es asimilada por el organismo y se excreta en forma de materia fecal. Una parte de la energía asimilada es utilizada posteriormente por el organismo para actividades de mantenimiento, como la respiración; esta energía se convierte en último término en calor y se pierde para el sistema. La fracción de energía que supera las puras necesidades de mantenimiento se convierte en el crecimiento de la población, ya sea en forma de un aumento del peso del cuerpo o bien en forma de una mayor fertilidad. La utilización de energía en una población de consumidores para el crecimiento o para un aumento en la reproducción se deno-

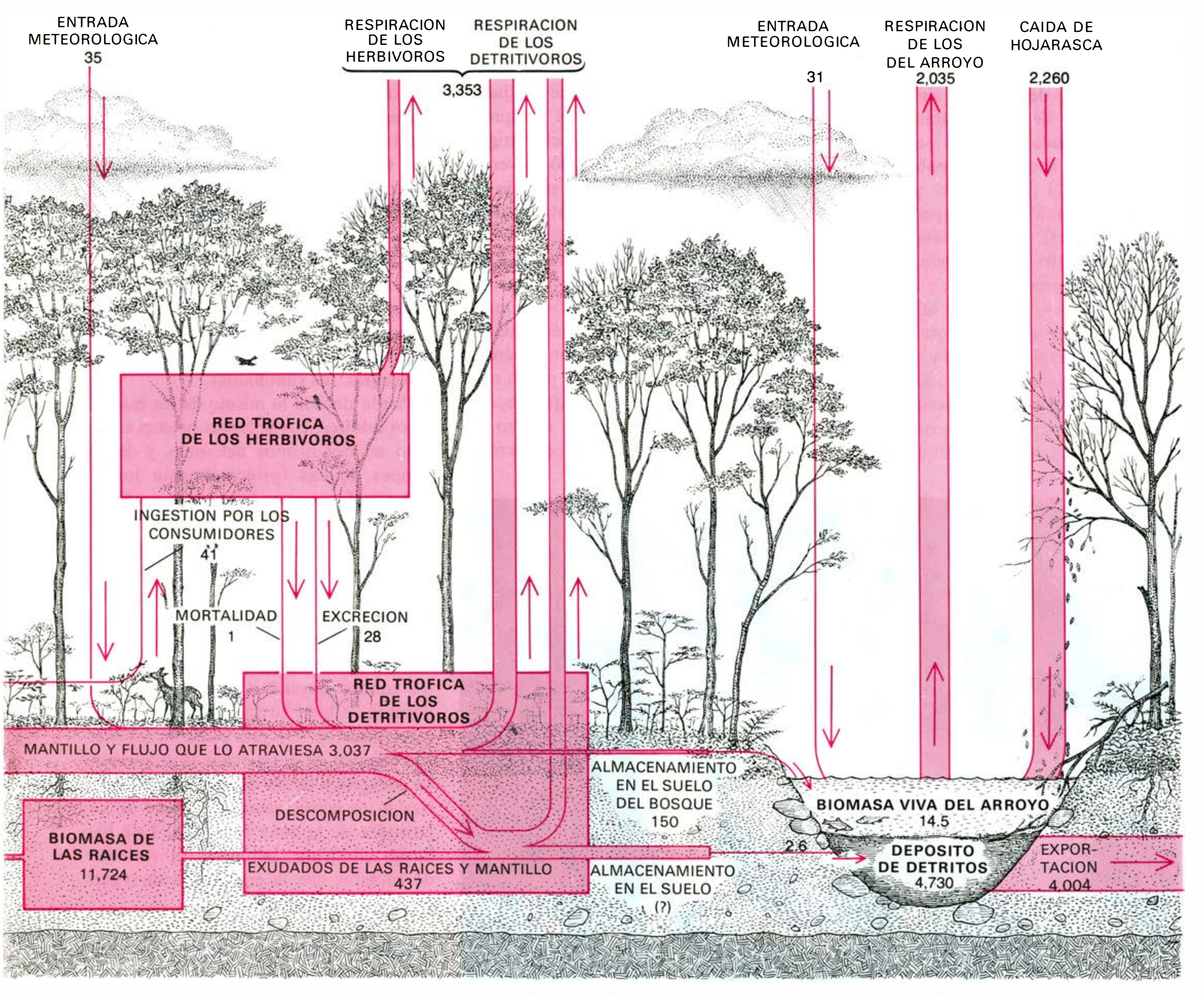


LA ENERGIA EN UN BOSQUE CADUCIFOLIO se distribuye de la forma que se indica en este diagrama de un ecosistema en el interior del Bosque Experimental del Hubbard Brook, en New Hampshire. El ecosistema es una cuenca que está limitada por divisorias colinares y que se asienta sobre un lecho rocoso impermeable. Las cifras representan entradas y salidas de energía en unidades de kilocalorías por metro cuadrado y por año; las energías relativas se indican mediante la anchura de los canales (la anchura del canal de la producción vegetal total se ha ampliado arbitrariamente en la llave). Las entradas de energía comprenden la cantidad de radiación solar que se convierte en materia orgánica mediante la fotosíntesis, y las entradas directas de materia orgánica arrastrada por la precipitación. Las salidas abarcan la fracción de energía solar que se refleja o que se pierde en forma de calor; la energía consumida por la evaporación de agua (evapotranspiración); la energía con-

mina producción secundaria. Cada especie de consumidor en el ecosistema difiere en su abundancia y en la eficiencia con la que utiliza energía. Aunque todavía no se dispone de información detallada sobre los requerimientos energéticos de los descomponedores, como bacterias y hongos, y de sus depredadores invertebrados poseemos cálculos para varias especies animales del ecosistema del Hubbard Brook.

Las orugas pueden consumir grandes cantidades de follaje, pero la tasa a la que extraen energía de su alimento es baja. Nuestras mediciones indican que *Heterocampa guttivita* asimila sólo alrededor de un 14 por ciento de la energía que hay en el tejido foliar que ingiere. El resto es excretado en forma de material fecal y cae al suelo del bosque, donde entra en la red trófica de los detritívoros. De la energía asimilada por la oruga, un 60 por ciento es utilizado para la respiración, y alrededor del 40 por ciento se incorpora a nuevo tejido. Esta producción de tejido forma una fracción de la base energética para gran variedad de consumidores, entre ellos depredadores invertebrados, aves y mamíferos, tanto de la red trófica de los detritívoros como de la de los herbívoros.

Las musarañas y las salamandras constituyen los dos grupos más importantes de animales vertebrados que capturan y comen los invertebrados del suelo del bosque. Hay dos especies de musarañas y cinco especies de salamandras en nuestro ecosistema, todas las cuales se alimentan principalmente de invertebrados. Puesto que las musarañas mantienen una temperatura corporal alta y relativamente constante, son activas durante todo el año, incluso bajo el manto invernal de nieve. Utilizan aproximadamente el 98 por ciento de su energía asimilada para la respiración y sólo un 2 por ciento para el crecimiento o para la reproducción. Con unos requerimientos metabólicos tan elevados, las musarañas deben comer a menudo para sobrevivir. Una musaraña colicorta *(Blarina brevicauda)* adulta, del Hubbard Brook, pesa unos 15 gramos y precisa unas 15,2 kilocalorías de energía en forma de alimento por día. Las



tenida en la materia orgánica que es transportada fuera del sistema por la corriente y la energía que se pierde en forma de calor a consecuencia de la respiración de los organismos en el ecosistema. Un exceso de entradas con respecto a las salidas produce un aumento en la cantidad total de energía almacenada en el ecosistema (la productividad neta del ecosistema). La mayor parte de la energía contenida en la producción anual de tejido vegetal entra en la cadena trófica de los detritívoros, ya sea directa o indirectamente, lo que produce una enorme acumulación progresiva de energía en y sobre el suelo del bosque. También se transfiere energía de los compartimientos de bosque del ecosistema al arroyo mediante filtración y caída de hojarasca. Puesto que la relación superficie de tierra a superficie del arroyo es de cerca de 570:1, pequeñas transferencias de energía desde cada metro cuadrado de tierra significan grandes entradas de energía en el arroyo. (El autor de este esquema es Tom Prentiss.)

presas de las musarañas (arañas, larvas de escarabajos y otros invertebrados del suelo) contienen por término medio alrededor de 0,11 kilocalorías por cada individuo presa, de modo que una musaraña típica necesitaría 138 de tales "comidas" al día. Suponiendo que la musaraña sea activa durante 12 horas del día, tendría que capturar y comer una presa cada cinco minutos aproximadamente. Estos valores oscilarán algo según la densidad y el tipo de presa, la temperatura y otras variables ambientales, pero está claro que queda poco tiempo en la vida de una musaraña para hacer algo que no sea comer.

En contraste con las musarañas, las salamandras son anfibios, animales de sangre fría que utilizan, en proporción, menos energía en la respiración, en particular durante los meses fríos del año, época en la que sus procesos metabólicos se enlentecen de manera significativa. La población de salamandras consume sólo aproximadamente un sexto de la energía que necesita la población de musarañas (1,1 kilocalorías por metro cuadrado y por año, en comparación con 7,1 kilocalorías), pero aproximadamente el 60 por ciento de su energía asimilada se dirige a la producción de nuevos individuos. De manera que las salamandras son muy eficientes a la hora de transformar la energía asimilada en un aumento de biomasa. Es interesante comprobar que musarañas y salamandras, todas ellas animales vertebrados que ocupan hábitats similares en el ecosistema, han desarrollado estrategias tan divergentes.

Las aves constituyen también un grupo importante de depredadores de insectos en las cuencas forestadas del Hubbard Brook. Su abundancia varía desde un máximo de 322 individuos por 10 hectáreas a mediados de verano, cuando los jóvenes acaban de abandonar el nido, hasta un mínimo de 14 por 10 hectáreas en pleno invierno. Combinando información acerca de las densidades de población y de las fluctuaciones de biomasa a lo largo del año con cálculos de laboratorio, sobre utilización de energía por unidad de masa, hemos estimado que la población de aves del Hubbard Brook asimila 5,2 kilocalorías por metro cuadrado y por año. Como ocurre con las musarañas, el 98 por ciento de esta energía se consume en la respiración, mientras que el dos por ciento restante se dedica a la producción de nuevo tejido.

La población de aves del bosque es muy móvil, y muchas especies migran hacia el sur antes de la estación invernal, transportando de este modo energía fuera del sistema. Nuestros datos indican que la biomasa de la población de aves en el ecosistema forestal permanece bastante constante de un año para otro, lo que sugiere que la producción de nuevo tejido (0,1 kilocalorías por metro cuadrado y por año) se equilibra con la pérdida de biomasa de aves causada por la mortalidad. Es difícil establecer qué fracción de esta mortalidad ocurre dentro del ecosistema y cuál fuera de él, pero en cualquier caso la cantidad de energía implicada es bastante reducida. Aun suponiendo que toda la mortalidad de las aves migradoras ocurra fuera del ecosistema, la máxima cantidad de energía perdida en la migración sería de 0,1 kilocalorías por metro cuadrado y por año, una fracción minúscula del flujo total de energía que atraviesa el sistema.

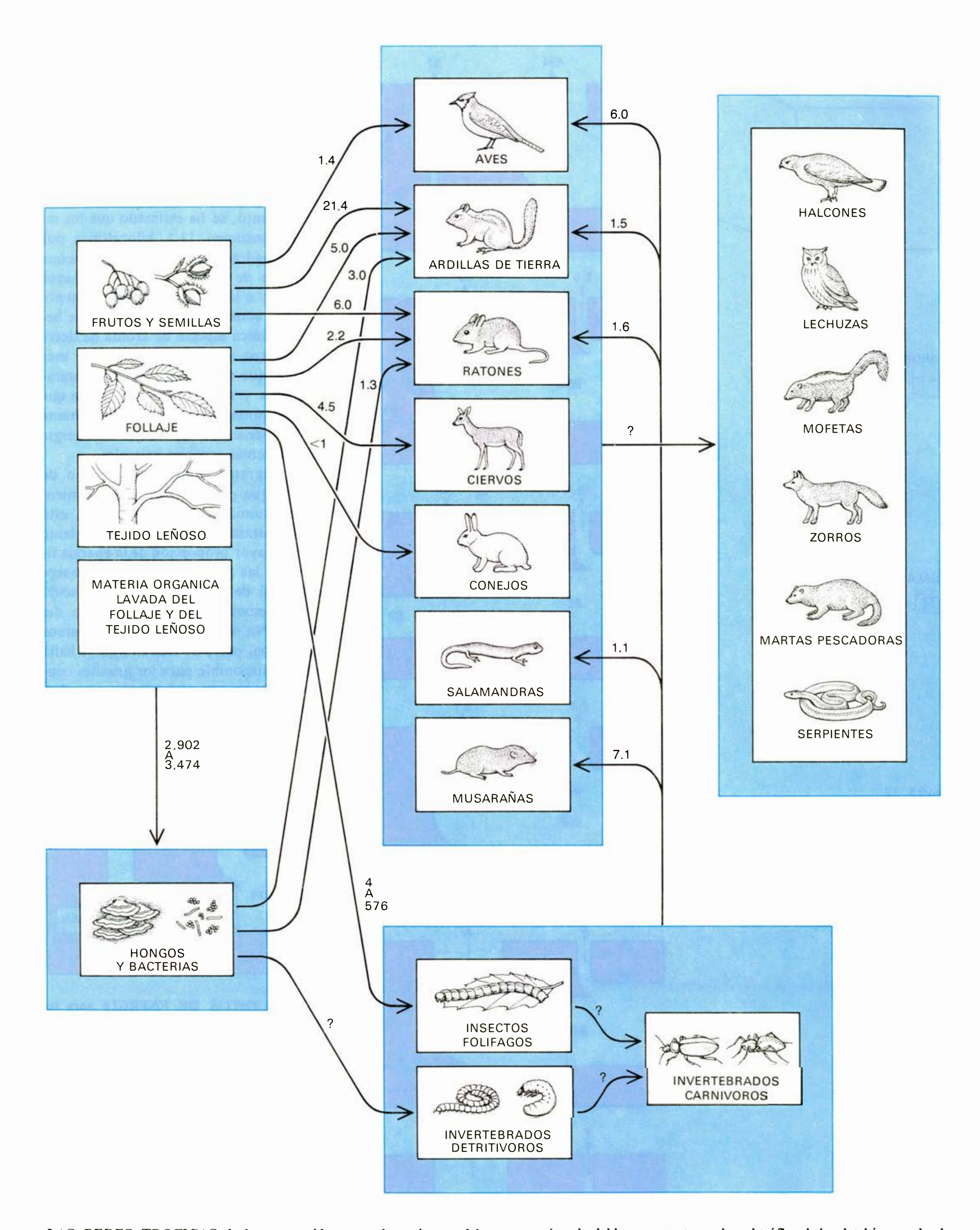
El 81 por ciento del alimento que ingieren las aves en el bosque es de origen animal, casi exclusivamente insectos; el resto es materia vegetal, principalmente frutos y semillas que se obtienen durante el otoño y el invierno. En los años en que las orugas folífagas son abundantes, estos insectos constituyen la mayor parte de la dieta de la población de aves, de modo que las aves están claramente ligadas a la red trófica de los herbívoros. Sin embargo, en los años en los que las orugas no abundan, hemos encontrado que muchas especies de aves forestales comen sobre todo insectos adultos, principalmente moscas, avispas y escarabajos, todos los cuales se hallan de algún modo relacionados con la red trófica de los detritívoros. Por ejemplo, las moscas adultas están representadas ampliamente por grupos cuyas larvas se alimentan de materia orgánica muerta en el suelo del bosque (las moscas comunes, los múscidos y los mosquitos de los hongos, los micetofílidos) o en los arroyos (las moscas negras, los simulíidos).

Puede decirse lo mismo de los escarabajos adultos, cuyas larvas predan sobre todo en organismos del suelo, y de las avispas adultas (principalmente los icneumónidos) cuyas larvas son parásitas de insectos de la red trófica de los detritívoros. Parece, por tanto, que aunque las aves explotan a los insectos folífagos cuando éstos existen, una fracción significativa de su dieta estival (y, en consecuencia, de su entrada total de energía) consiste en insectos adultos cuyas larvas se alimentan de detritos, o bien en los depredadores de estos insectos. De esta manera, las aves son capaces de aprovechar la gran reserva energética de los detritos que caracteriza a los ecosistemas de bosque de las latitudes de la zona templada. Este hallazgo confirma, asimismo, la variedad de caminos que la energía puede seguir en un ecosistema de este tipo.

La fauna de roedores del bosque del Hubbard Brook está constituida principalmente por la ardilla de tierra oriental *(Tamias striatus)* y tres especies de ratones: el ratón de patas blancas *(Peromyscus maniculatus)*, el ratón saltador de bosque *(Napaeozapus insignis)* y el topino de Gapper *(Clethrionamys gapperi)*. Estos roedores obtienen directamente



ESTA LARVA DE LA POLILLA *Heterocampa guttivita* es una oruga defoliadora, indígena de la región del Hubbard Brook. En los años en los que sobreabunda puede consumir más del 40 por ciento de la producción anual de follaje. Las orugas asimilan sólo un 14 por ciento de la energía del tejido de las hojas ingeridas; el resto se excreta y entra en la red trófica de los detritívoros.



LAS REDES TROFICAS de los consumidores en el ecosistema del Hubbard Brook se esquematizan aquí, junto con la tasa de consumo de organismos en cada población. Todos los datos numéricos corresponden a kilocalorías por metro cuadrado y por año. El consumo de hojas por parte de los insectos herbívoros varía mucho de un año para otro, y puede incidir ampliamente en la utilización de energía por parte de otros consumidores. Casi todos los años, una gran cantidad de materia orgánica cae al suelo del bosque y entra en la red trófica de los detritívoros, donde es utilizada por hongos, bacterias y algunos invertebrados. Estos organismos son el alimento de invertebrados carnívoros, salamandras, musarañas y de algunos animales asociados principalmente a la red trófica de los herbívoros. Las aves participan en ésta al comer bayas y orugas, pero pueden aprovechar el gran acúmulo de energía de los detritos característicos de los ecosistemas de los bosques caducifolios septentrionales.

ORUGAS CON DENSIDAD ELEVADA 160.4

494 51 576 82 31

ARDILLAS DE TIERRA .21

5.5 25 31 25.5 .5

SALAMANDRAS .20

.2 .34 1.06 .86 .52

AVES .12

2.2 5.1 7.4 5.2 .1

RATONES .08

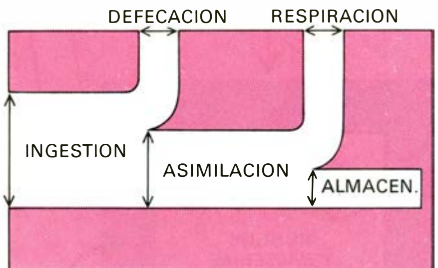
1.9 9.0 11.1 9.2 .2

MUSARAÑAS .02

.7 6.3 7.1 6.4 .1

energía de los productores primarios (hojas, brotes, semillas y yemas), de otros consumidores (invertebrados del mantillo del bosque y, ocasionalmente, huevos y polluelos de aves) y de descomponedores (hongos); las presas incluyen, pues, miembros de la red trófica de los herbívoros y de la de los detritívoros. En conjunto, se ha estimado que los ratones consumen 11,1 kilocalorías por metro cuadrado y por año, mientras que la ardilla de tierra, debido a su mayor tamaño y a la mayor densidad de su población, consume casi el triple. De hecho, la única especie de ardilla de tierra de nuestro ecosistema consume más energía que todos los ratones, musarañas, aves y salamandras juntos, lo que la hace potencialmente más importante en la estructura del flujo de energía que las demás especies animales.

El diagrama completo del flujo de energía que pasa por los compartimientos consumidores del ecosistema está ahora bastante bien trazado. Es evidente que la mayor proporción de la energía fijada por las plantas en este bosque septentrional de caducifolios es conducida a los descomponedores del suelo del bosque. No obstante, como indicábamos más arriba, parte de esta energía se halla todavía disponible para los grandes consumidores, puesto que mantiene muchas poblaciones de insectívoros: invertebrados, anfibios y mamíferos. Incluso las aves de la bóveda arbórea consiguen ex-



PRESUPUESTOS DE ENERGIA para los principales organismos consumidores que viven sobre el suelo en el bosque del Hubbard Brook, representados de manera esquemática. El tamaño de los cubos es proporcional a la biomasa de cada población a mediados de verano, cuando es máxima. Los canales que atraviesan los compartimientos representan la eficiencia en la utilización de energía por cada organismo. La energía ingerida en forma de materia orgánica pasa a través del tubo digestivo y es excretada o bien es asimilada por el organismo. La energía asimilada se gasta luego en la respiración o se invierte en la producción de nuevos tejidos, a través del crecimiento o de la reproducción. Los animales de sangre fría, como los insectos y las salamandras, son más eficientes en invertir energía en un aumento de biomasa, mientras que los vertebrados de sangre caliente, como los mamíferos y las aves, utilizan cerca del 98 por ciento de su energía asimilada para respirar, y sólo un dos por ciento para el crecimiento.

plotar la reserva energética de los detritos al alimentarse de insectos adultos cuyas larvas viven de detritos.

Aunque hemos progresado mucho en el camino hacia la comprensión de las relaciones energéticas y, por tanto, de las relaciones funcionales dentro del ecosistema de bosque, las relaciones entre los compartimientos de los consumidores y los factores que regulan el flujo de energía en el interior del sistema sólo ahora empiezan a dilucidarse en su total complejidad. Por ejemplo, ¿cuál es la importancia de los insectos folífagos al afectar la tasa de fijación de energía por las plantas? Como ya hemos mencionado, en determinados años los insectos defoliadores consumen hasta el 44 por ciento del tejido de las hojas verdes del ecosistema, lo que sugiere que representan una incidencia importante en la producción primaria, pero todavía deben determinarse los efectos a largo plazo de la defoliación realizada por los insectos en la productividad vegetal. Asimismo, ¿qué influencia tienen los comedores de semillas (aves, mamíferos e insectos) en la reproducción de los árboles del bosque? ¿Y cómo regulan los organismos de la red trófica de los detritívoros la tasa en el ciclado de los nutrientes y, en consecuencia, la productividad del bosque? Puesto que muchos organismos del suelo se alimentan de tejido radicular vivo, pueden actuar como consumidores en la red trófica de los herbívoros y así afectar a las plantas en su capacidad de absorber nutrientes y de fijar energía.

Habrá que responder a éstas y otras preguntas para que podamos garantizar un conocimiento completo de los efectos reguladores de los consumidores en el funcionamiento de los ecosistemas forestales. De la consideración de la influencia de las actividades del hombre en los sistemas naturales surgen nuevos interrogantes. Por ejemplo, se sabe que los contaminantes atmosféricos, como el ozono y los compuestos ácidos disueltos en la precipitación, tienen efectos adversos sobre la vegetación, pero no se sabe casi nada de sus efectos sobre los organismos consumidores o descomponedores. ¿Acaso los consumidores representan un mayor impacto sobre el flujo de energía cuando la comunidad vegetal se halla sometida a estrés por un tóxico externo? Estas preguntas no resultan de la pura especulación. Es esencial una profunda comprensión de los mecanismos reguladores naturales que influyen en el funcionamiento de los ecosistemas si han de desarrollarse planes racionales para la gestión de los bosques.

Cómo fabrican ATP las células

La teoría predominante es la «quimiosmótica». La luz o la oxidación conducen protones a través de la membrana; al retornar éstos por entre un conjunto de enzimas, se forma el compuesto ATP, rico en energía

Peter C. Hinkle y Richard E. McCarty

La energía adquirida por las células vivas se conserva de manera aprovechable principalmente en forma de moléculas de adenosín trifosfato, que se abrevia por ATP. Tanto si la energía proviene en último extremo de la luz como si procede de la oxidación de compuestos orgánicos, se invierte en su mayor parte en la elaboración de ATP; éste pasa a ser entonces el «fluido energético» que pondrá en marcha las demás funciones de la célula.

Hoy se conoce bastante bien el conjunto de reacciones químicas que conducen a la síntesis de ATP. En los cloroplastos de las células de plantas verdes se extraen, a partir del agua, átomos de hidrógeno, o electrones y protones. Los protones quedan en disolución y los electrones son conducidos por la energía de la luz a través de una serie de moléculas transferidoras. Eventualmente, los electrones y los protones se combinan con dióxido de carbono formando moléculas orgánicas. En las mitocondrias de todas las células con núcleo, los electrones cedidos por moléculas orgánicas atraviesan una cadena similar de moléculas transferidoras y, finalmente, son aceptados por el oxígeno, formando agua. Muchos de los estados intermedios de estas transformaciones energéticas han sido desentrañados, pero quedaba una fase crucial por desvelar. No estaba claro cómo la transferencia de electrones a través de la serie de moléculas transferidoras se acoplaba con la síntesis de ATP. La transferencia de electrones proporciona suficiente energía para la formación de ATP, pero no ha resultado tarea fácil describir el mecanismo de su síntesis.

Peter Mitchell, de los Glynn Research Laboratories, de Inglaterra, propuso, en 1961, un hipotético mecanismo del acoplamiento entre la transferencia de electrones y la síntesis de ATP. Mitchell sugirió que el flujo de electrones a través del sistema de moléculas transferidoras conducía hidrogeniones cargados positivamente, o protones, a través de las membranas de cloroplastos, mitocondrias y células bacterianas. A consecuencia de ello se creaba un gradiente electroquímico de protones a través de la membrana. El gradiente constaba de dos componentes: una diferencia en la concentración de hidrogeniones, o *p*H, y una diferencia de potencial eléctrico. La síntesis de ATP se llevaba a cabo por un flujo de protones en sentido inverso, a favor del gradiente. La propuesta de Mitchell se denominó teoría quimiosmótica. Durante los últimos quince años, los trabajos de Mitchell y de su colega Jennifer M. Moyle, así como los de otros muchos investigadores, han mostrado que los postulados básicos de la teoría quimiosmótica son, en su mayoría, ciertos, aún cuando algunos detalles siguen creando controversias.

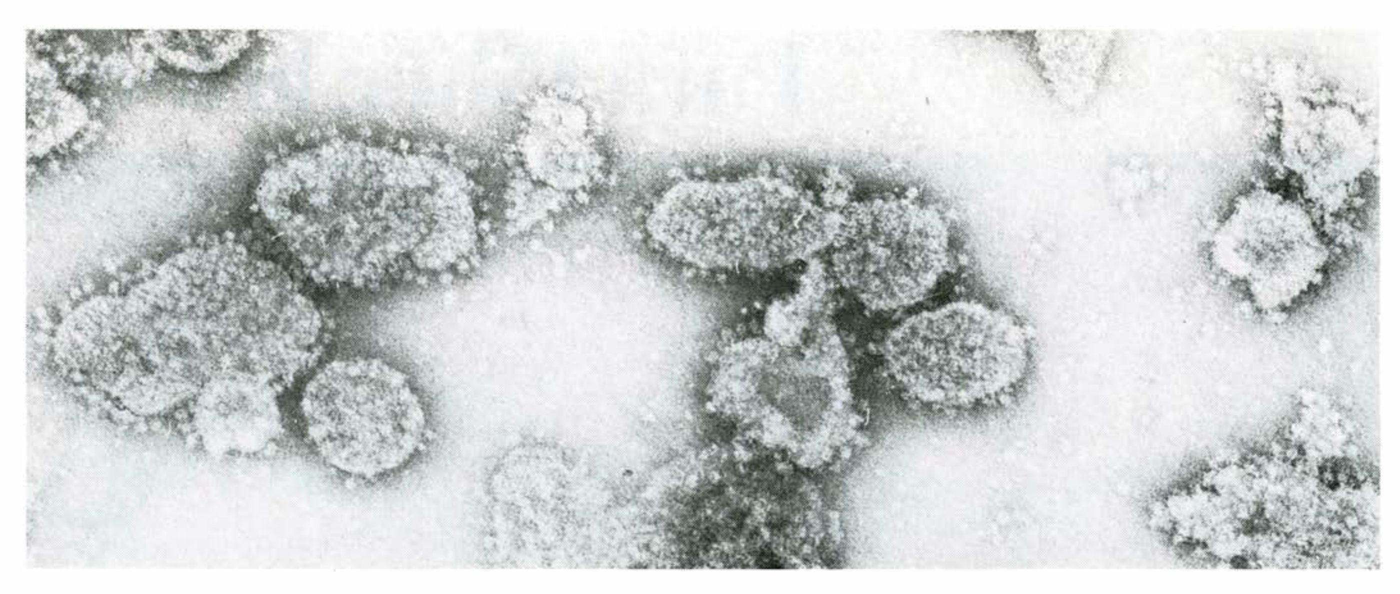
En la teoría quimiosmótica no se postula la existencia de ninguna maquinaria molecular en la que el flujo de energía procedente de la luz o de la oxidación se acople a la síntesis de ATP. En su lugar, el papel crucial lo desempeña una membrana que separa una región de otra. La membrana proporciona algo más que protección, confinamiento y un medio interno controlado, ya que la disposición asimétrica de las moléculas transferidoras a través de la membrana permite el establecimiento del gradiente de protones. Las mismas propiedades topológicas de la membrana ayudan grandemente a mantener el gradiente de protones: debe formar una envoltura cerrada. Así, en el metabolismo energético, la distinción más importante se presenta entre el interior y el exterior.

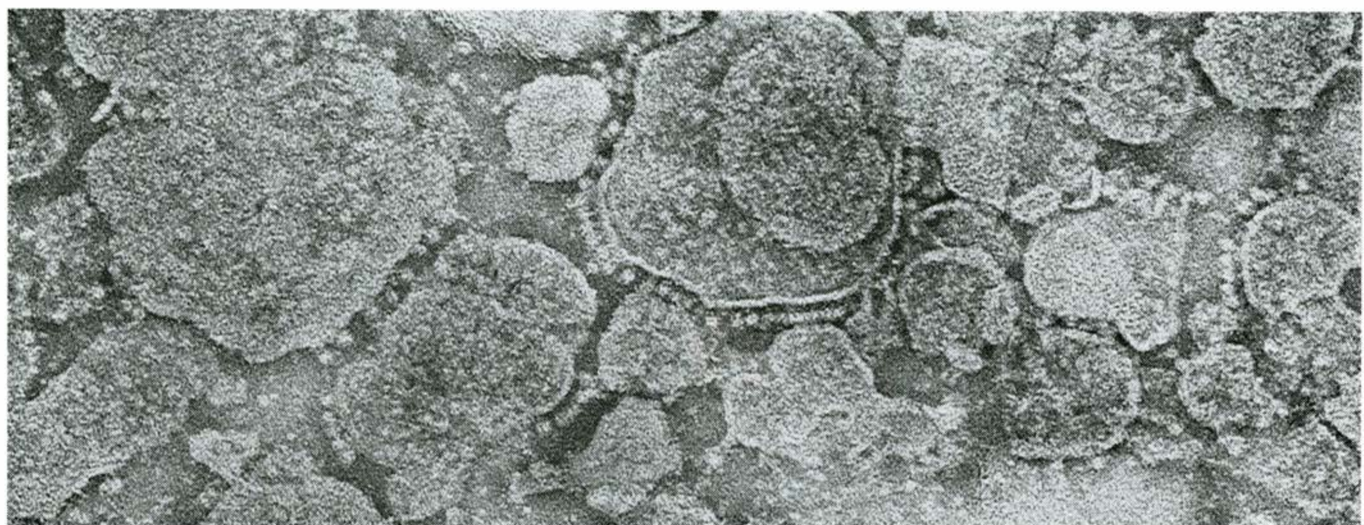
Un concepto esencial para la comprensión de la síntesis de ATP es el de energía libre, que mide la cantidad de energía de un sistema químico disponible para efectuar un trabajo. Cualquier reacción química, tanto si tiene lugar en la célula viva como si no, únicamente puede discurrir en el sentido de una disminución de la energía libre.

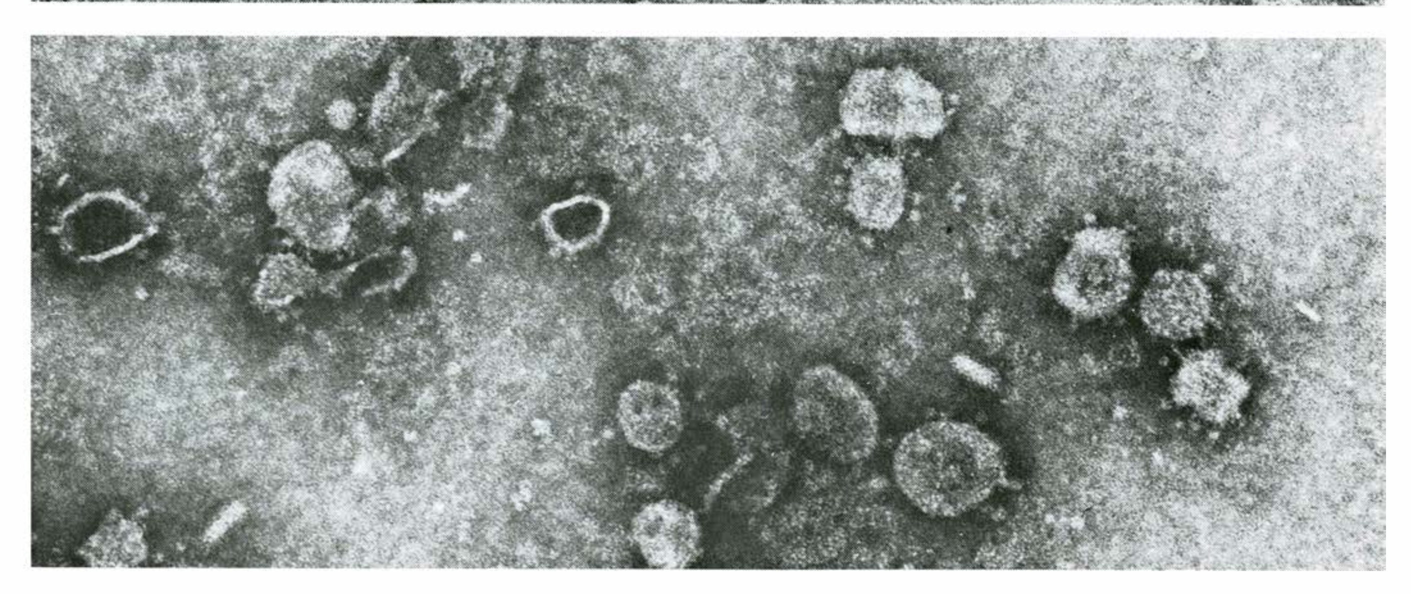
La reacción de los hidratos de carbono con el oxígeno (para producir dióxido de carbono y agua) libera una gran cantidad de energía libre. Por tanto, la reacción, que tiene lugar en el proceso fisiológico denominado respiración, está termodinámicamente favorecida. En la fotosíntesis, la misma reacción global se produce en sentido contrario: el dióxido de carbono se combina con agua para formar hidratos de carbono y oxígeno molecular (O_2). Esta reacción, si se considera aisladamente, correspondería a un aumento de la energía libre, por lo que sólo puede tener lugar si hay un aporte de energía procedente de una fuente exterior. La energía, naturalmente, proviene de la luz solar; deben absorberse varios fotones, o cuantos de luz, por cada molécula de dióxido de carbono convertida en hidrato de carbono. Los fotones son absorbidos por moléculas de pigmentos que, en consecuencia, quedan en un estado de excitación con elevada energía libre. La reacción global, incluyendo los pigmentos excitados entre las moléculas reaccionantes, discurre, como requieren las leyes de la termodinámica, en el sentido de una disminución de la energía libre.

La síntesis de ATP está afectada por restricciones energéticas del mismo tipo. Una molécula de ATP consta de una base orgánica que contiene nitrógeno (adenina), un azúcar (ribosa) y una cadena de tres grupos fosfato. Es un ejemplar de las moléculas denominadas nucleótidos. En la mayoría de las reacciones en las que el ATP actúa de fuente de energía sólo está implicado el grupo fosfato terminal. El ATP se forma enlazando un tercer grupo fosfato al adenosín difosfato (ADP), con eliminación de agua. La reacción no tiene lugar espontáneamente, sino que debe suministrarse una elevada cantidad de energía. Mucha de esta energía puede recuperarse luego en la reacción inversa, en la que el ATP se disocia en ADP y fosfato inorgánico. Al ceder el grupo fosfato terminal a otras moléculas, el ATP da lugar a especies fosforiladas de alta energía libre que pueden intervenir en reacciones que de otra forma no podrían tener lugar.

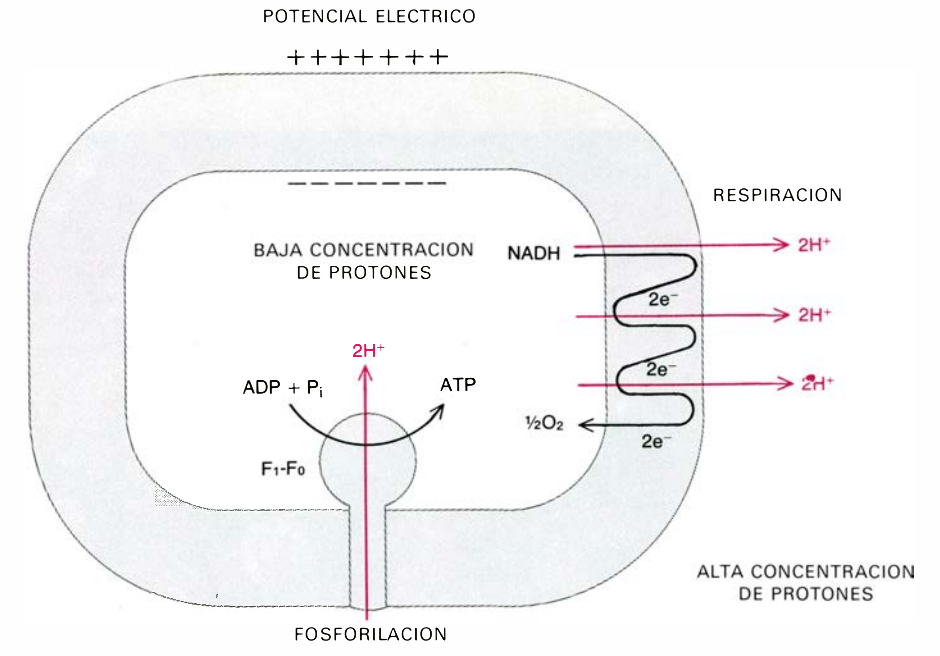
Las transacciones energéticas bioquímicas se miden convencionalmente tomando



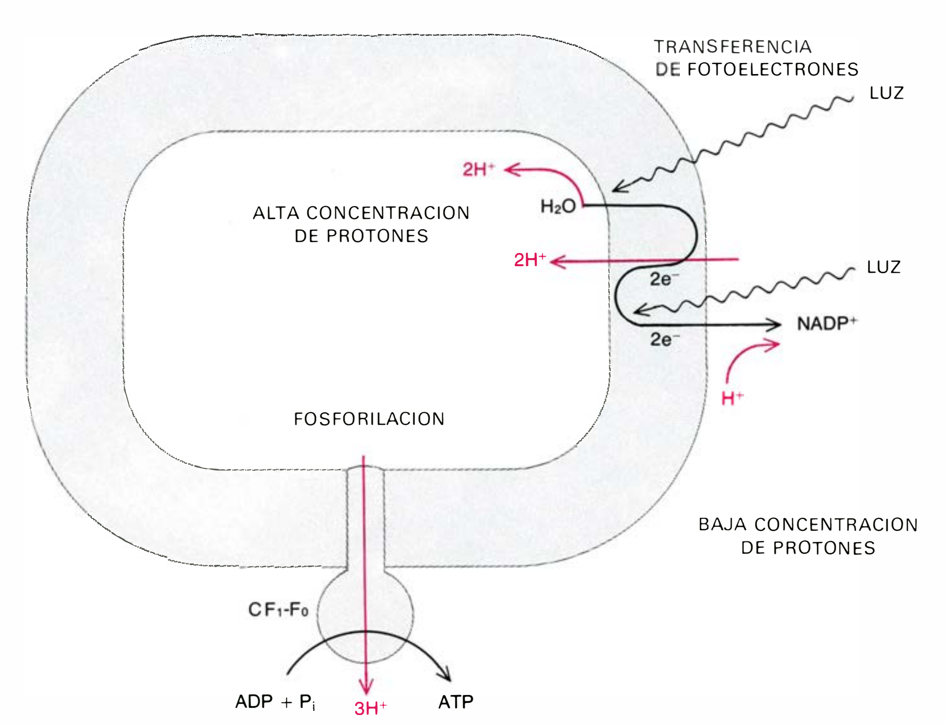




ESTAS VESICULAS DE MEMBRANA muestran unos «botones» que han sido identificados como parte de un complejo enzimático capaz de sintetizar adenosín-trifosfato, o ATP. Las vesículas de la parte superior fueron preparadas a partir de membranas internas de los orgánulos subcelulares denominados mitocondrias; las del centro provienen de la membrana interna de cloroplastos, que son los orgánulos fotosintéticos de las células de plantas verdes; las vesículas de la parte inferior proceden de membranas de la bacteria *Escherichia coli.* Todos estos sistemas de membranas son capaces de sintetizar ATP. La similaridad morfológica sugiere que utilizan mecanismos análogos. Las protuberancias constan de varias subunidades proteicas, conocidas en conjunto como F_1; están unidas a una proteína de la membrana denominada F_0. En las vesículas mostradas aquí, la proteína F_1 sobresale de la superficie exterior de la membrana, pero ésta es sólo su orientación natural en los cloroplastos. En las mitocondrias y las bacterias la proteína se halla habitualmente en el interior; en este caso está exteriormente porque las membranas disgregadas forman vesículas vueltas del revés. Las vesículas mitocondriales y las de cloroplastos han sido preparadas por los autores y las bacterianas por E. Hertzberg. Las tres microfotografías de las vesículas de membrana aquí presentadas han sido obtenidas por John N. Telford, de la Universidad de Cornell.



LA FOSFORILACION OXIDATIVA acopla la liberación de energía procedente de la oxidación de moléculas derivadas de hidratos de carbono y de lípidos con la síntesis de ATP. En las células nucleadas, la fosforilación oxidativa tienen lugar sólo en las mitocondrias. Los hechos fundamentales del proceso mitocondrial se representan aquí tal como los describe la teoría quimiosmótica. Los electrones y los protones (o átomos de hidrógeno) procedentes de los hidratos de carbono o de los lípidos son conducidos por moléculas del transferidor de hidrógeno NADH a un sistema de enzimas que se halla localizado en la membrana mitocondrial. Durante la respiración, parejas de electrones atraviesan la membrana por tres veces, transportando en cada una de ellas dos protones desde el interior al exterior de la mitocondria. Se forma así un gradiente en la concentración de protones y un potencial eléctrico que tiende a introducir de nuevo los protones a través de la membrana. La energía del gradiente dirige el proceso de síntesis de ATP. Por cada dos protones que atraviesan el complejo F_1-F_0 se forma una molécula de ATP a partir de ADP y de fosfato inorgánico (P_i).



LA FOSFORILACION FOTOSINTETICA de los cloroplastos obtiene de la luz la energía necesaria para la formación de ATP. Al igual que en la fosforilación oxidativa, se transportan iones de hidrógeno a través de la membrana para crear un gradiente de protones, y se sintetiza ATP al retornar los protones a través de la membrana a favor del gradiente. No obstante, en los cloroplastos, el sentido del flujo de los protones está invertido: el movimiento de los electrones, dirigido por la luz, bombea los protones hacia dentro, acidificando el interior, y la fosforilación depende de un flujo hacia el exterior. Más aún, la estequiometría, o proporción de reactivos e iones, es diferente de la observada en las mitocondrias. Cada par de electrones sólo cruza la membrana dos veces, transfiriendo sólo cuatro protones, y por cada molécula de ATP formada deben pasar tres protones a través del complejo enzimático, designado CF_1-F_0.

como unidad la kilocaloría por mol. Una kilocaloría es la cantidad de energía necesaria para elevar en un grado Celsius la temperatura de un kilogramo de agua. Un mol es la cantidad de una substancia, expresada en gramos, numéricamente igual a su peso molecular; un mol de azúcar glucosa, por ejemplo, pesa 180 gramos. La oxidación completa de la glucosa proporciona alrededor de 700 kilocalorías por mol. Por tanto, cada mol de glucosa oxidado produce energía suficiente para la síntesis de alrededor de 46 moles de ATP. En realidad, por cada mol de glucosa oxidado no se forman más de 36 moles de ATP, y a veces tan sólo 25. El acoplamiento de estas reacciones está claramente permitido desde el punto de vista termodinámico.

Si un proceso químico conduce a un estado de menor energía libre la reacción puede tener lugar, pero no debe darse por sentado que necesariamente será así. La glucosa y el oxígeno, por ejemplo, son bastante estables a temperatura ambiente; reaccionan (la glucosa se quema) únicamente si son calentados. El calor aplicado representa una energía de activación, que se recupera junto con las 700 kilocalorías por mol desarrolladas por la oxidación. En los sistemas biológicos, la necesidad de una energía de activación la reducen las proteínas catalíticamente activas que reciben el nombre de enzimas. Dado que en casi la totalidad de las reacciones bioquímicas intervienen enzimas, pueden producirse a temperaturas y presiones fisiológicas, y quedan confinadas a vías metabólicas determinadas. Es importante hacer notar que los enzimas no afectan el sentido en el que discurre una reacción sino sólo su velocidad. La acción de un enzima es, en cierto modo, análoga a la de un lubricante: no puede impulsar la reacción cuesta arriba, pero puede hacerla ir más aprisa cuesta abajo.

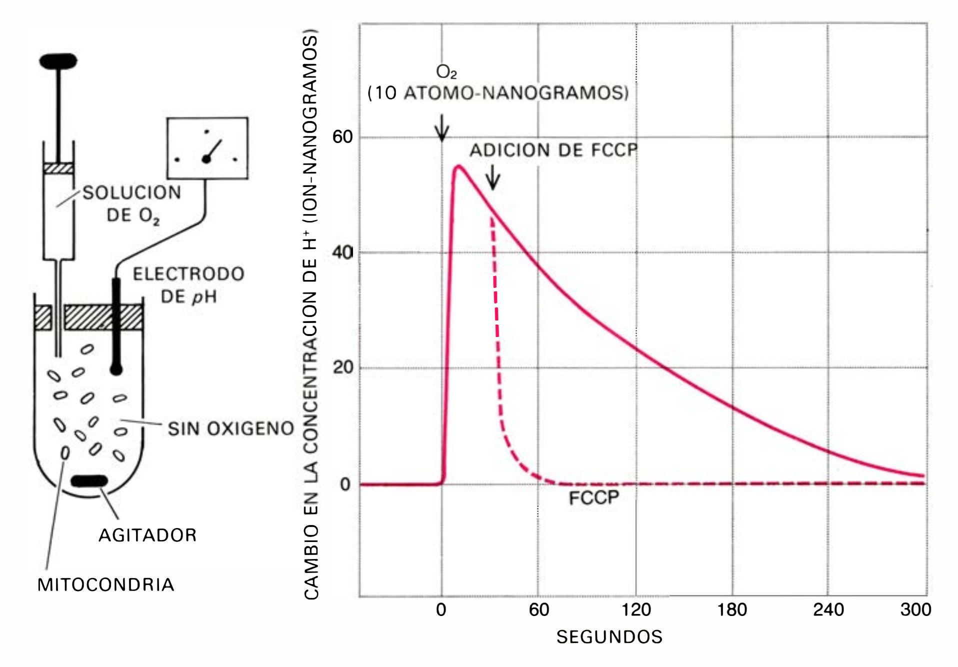
En el metabolismo de oxidación casi el 95 por ciento de la energía almacenada en forma de ATP procede de la transferencia de electrones desde los hidratos de carbono, u otros substratos, hasta el oxígeno. En la fotosíntesis toda la energía se deriva del transporte de electrones en el sentido contrario, del agua al dióxido de carbono. La teoría quimiosmótica se circunscribe a las etapas del metabolismo que comienzan con transferencia de electrones y finalizan con síntesis de ATP. En las mitocondrias y en bacterias, estos procesos reciben el nombre de fosforilación oxidativa; en los cloroplastos y en bacterias fotosintéticas se llaman fosforilación fotosintética. Quedan por aclarar muchos puntos de ambos sistemas. Presentaremos aquí la teoría sobre la fosforilación que actualmente parece estar más de acuerdo con las observaciones experimentales.

Globalmente, los fenómenos químicos de oxidación y fotosíntesis consisten en una transferencia de hidrógeno, pero no es necesario que en cada uno de los estados del proceso se produzca el transporte de átomos de hidrógeno completos. Es más, en la teoría quimiosmótica los transferidores de hidrógeno alternan con moléculas que únicamente conducen electrones. El empleo de transferidores de electrones es posible, ya que los protones son solubles en agua y en el medio acuoso de la célula, mientras que los electrones no lo son. Al interactuar una molécula que porta un átomo de hidrógeno completo con otra molécula que únicamente acepta electrones, se libera el protón a la solución. Si luego un transferidor de electrones cede su electrón a un transferidor de hidrógeno, el átomo de hidrógeno queda reconstituido al ser captado un protón del medio.

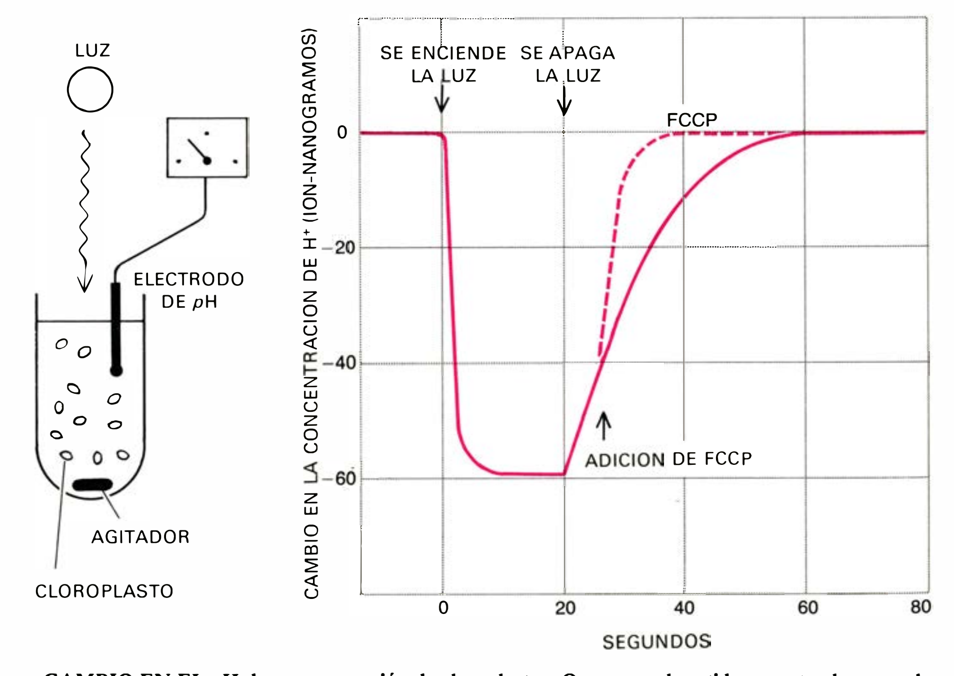
La transferencia de un electrón o un átomo de hidrógeno de una molécula a otra recibe el nombre de reacción de oxidación-reducción. La molécula que cede el electrón o el hidrógeno se dice que ha sido oxidada por la molécula que lo capta; a la inversa, se dice que el aceptor de electrones o de hidrógeno ha sido reducido. Siempre que una substancia sea oxidada, otra debe ser reducida. Es por ello por lo que el término oxidación puede crear a veces confusiones. El oxígeno es, desde luego, un agente muy oxidante, pero el proceso químico de la oxidación es general, y puede tener lugar en ausencia de oxígeno. Cualquier aceptor de electrones puede considerarse, con toda propiedad, oxidante.

En las mitocondrias, el hidrógeno es extraído a partir de los hidratos de carbono a través de la complicada serie de transformaciones que constituyen el llamado ciclo del ácido cítrico. Los detalles del ciclo no nos interesan en este momento; baste indicar que su efecto global es la degradación de la cadena carbonada de la glucosa a dióxido de carbono y el aporte de los átomos de hidrógeno liberados a la molécula de nicotinamida-adenín-dinucleótido, o NAD^+. El signo positivo se ha incluido en la abreviatura para indicar que la molécula tiene ordinariamente una carga eléctrica positiva. Cada molécula de NAD^+ acepta dos electrones y un protón. El protón y uno de los electrones se unen a un átomo de carbono de la molécula de NAD^+; el otro electrón neutraliza la carga positiva. Esta forma reducida de NAD^+ se designa NADH. El NADH es el principal intermediario entre el ciclo del ácido cítrico y los enzimas de la membrana interior de las mitocondrias que, eventualmente, suministran electrones al oxígeno, con lo que se forma agua. En conjunto, el proceso se llama respiración. Los transportadores de hidrógeno y los transportadores de electrones de la membrana componen la cadena respiratoria.

Según la versión de la teoría quimiosmótica sobre la respiración mitocondrial que en la actualidad parece más plausible, cada par de electrones transferidos del NADH al oxígeno dan lugar a la salida, a través de la membrana, de seis protones. La relación se expresa en términos de pares de electrones, ya que los electrones aparecen en parejas tanto al comienzo de la cadena respiratoria (en el NADH) como en su final (donde reducen un átomo de oxígeno a agua). A lo largo de la cadena respiratoria, los electrones son transporta-



TRANSPORTE DE PROTONES por mitocondrias activas. Se manifiesta aquí como un cambio en la concentración de hidrogeniones en el medio que rodea los orgánulos. Las mitocondrias se suspendieron en un medio que contenía una fuente de electrones para la respiración, pero del que se había extraído el oxígeno. Cuando se añadía una pequeña cantidad de oxígeno a la suspensión, el *p*H del medio descendía claramente, lo que correspondía a un incremento en la concentración de protones. Al agotarse el oxígeno, los protones retornaban lentamente a las mitocondrias. El FCCP, una molécula que hace a las membranas permeables a los protones, disipaba el gradiente de protones.



CAMBIO EN EL *p*H de una suspensión de cloroplastos. Ocurre en el sentido opuesto al que se observa con las mitocondrias. Los cloroplastos se suspendieron en un medio que contenía un aceptor de átomos de hidrógeno y se mantuvieron en la obscuridad. Al iluminar la suspensión, la concentración de hidrogeniones descendió (indicando que los protones estaban siendo captados por los orgánulos), y al apagar la luz el gradiente decaía lentamente. De nuevo aquí, la molécula de FCCP aceleraba la filtración de los protones, que, en este caso particular, era desde el orgánulo al medio.

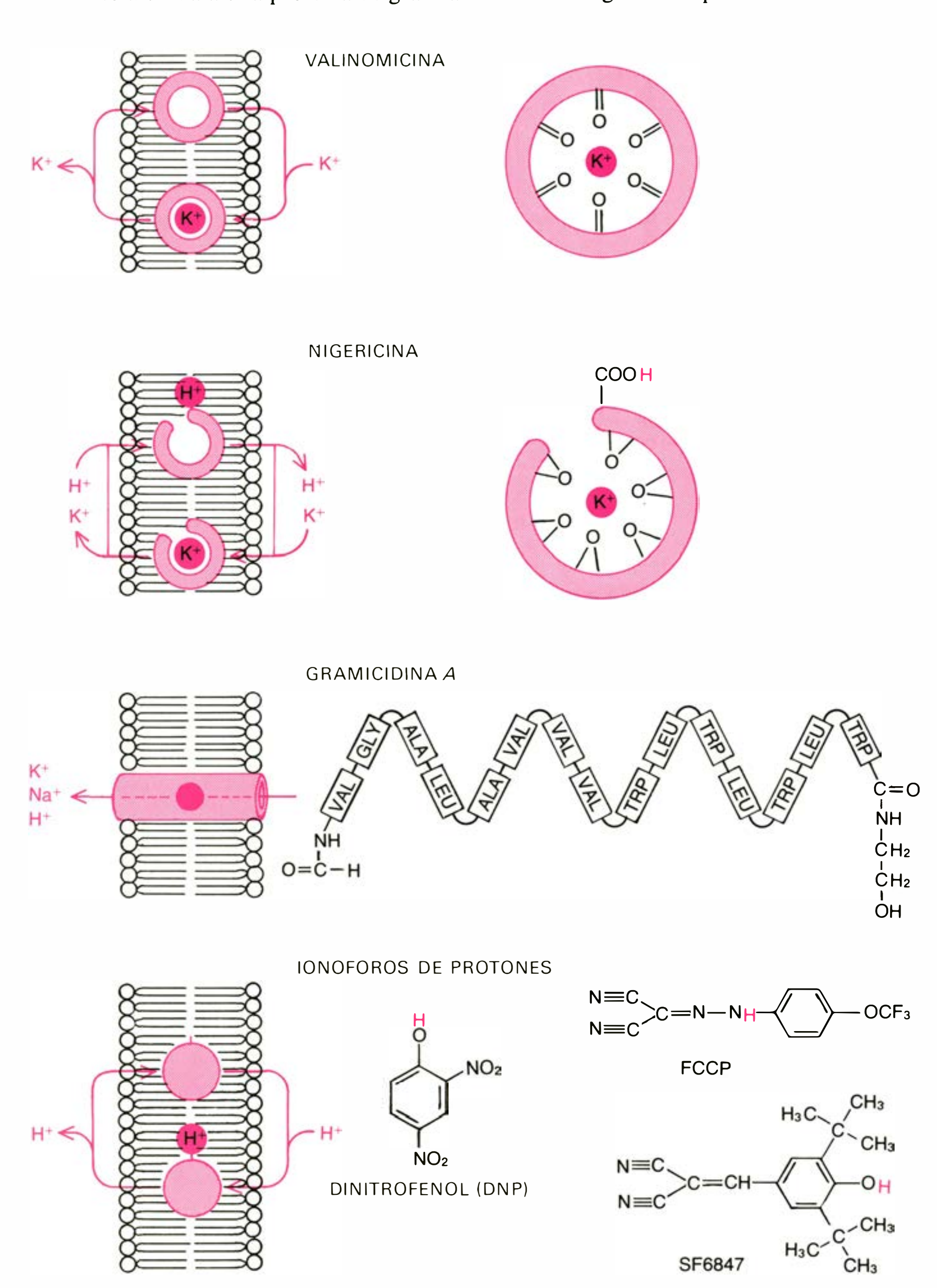
dos individualmente por ciertos transferidores y en parejas por otros.

El NADH cede sus dos electrones y un protón a un grupo transferidor denominado flavín-mononucleótido, o FMN. En este proceso el NADH es oxidado, es decir, retorna a la forma NAD^+, y el FMN, habiendo aceptado dos electrones y un protón, capta un protón adicional del medio interno delimitado por la membrana, con lo que es reducido a $FMNH_2$. La molécula de FMN está unida a una proteína de gran tamaño incluida en la membrana mitocondrial, y que, probablemente, la atraviesa por completo.

Mediante un mecanismo todavía desconocido, el $FMNH_2$ transfiere los dos átomos de hidrógeno desde la superficie interior de la membrana hasta la externa. Una vez allí los átomos son ionizados y los protones liberados al medio extramitocondrial, con lo que los dos primeros protones han sido ya transportados a través de la membrana. Según un hipotético mecanismo de transporte, los protones serían liberados por el grupo flavínico en el interior de la membrana y alcanzarían el medio externo a través de un canal de la molécula de proteína.

Al ser liberados los dos protones por el FMN, los dos electrones son transferidos a otras proteínas que, debido a los elementos asociados a ellas, reciben el nombre de ferroproteínas sulfuradas, o FeS. Tras ceder dos protones y dos electrones, el $FMNH_2$ retorna a su forma original FMN, y puede de nuevo ser reducido por el NADH.

A diferencia del FMN, las ferroproteínas sulfuradas únicamente transfieren electrones, no átomos completos de hidrógeno, y, es más, transportan los electrones de uno en uno y no a pares. Hay diversos tipos de ferroproteínas sulfuradas en este tramo de la cadena respiratoria; deben estar dispuestas a través de la membrana de forma que puedan transferir los electrones de nuevo al interior. Seguidamente, los electrones entran en la parte más complicada de la cadena respiratoria, en la que la teoría quimiosmótica es más especulativa. Mitchell ha propuesto recientemente un mecanismo para la transferencia de electrones en esta sección de la cadena, que expondremos, aunque de momento falta el apoyo experimental que corrobore esta suposición.

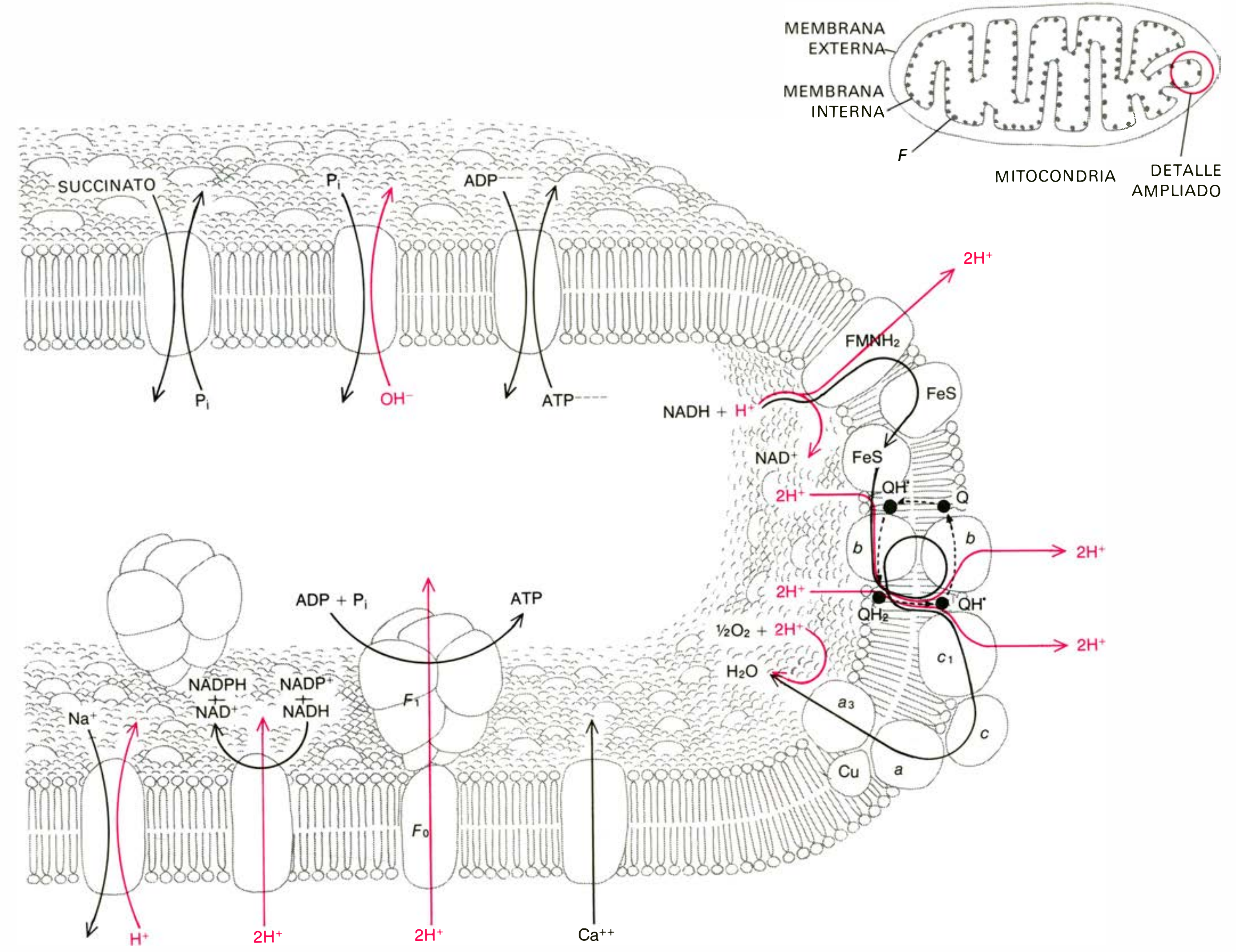


IONOFOROS son aquellas moléculas que transportan iones a través de las membranas. Se trata de unas herramientas esenciales en el estudio de los mecanismos quimiosmóticos. La mayoría de los iones simples son solubles en agua, pero no en la matriz lipídica de las membranas, de forma que las membranas se muestran impermeables a los iones. El ionóforo valinomicina es una molécula cíclica soluble en las membranas y que transporta selectivamente iones de potasio (K^+). La nigericina transporta también potasio, pero sólo intercambiándolo por protones; es insoluble en la membrana a menos que un ion o el otro estén ligados. La gramicidina *A* forma un poro en la membrana, compuesto de dos moléculas helicoidales, que es permeable a iones con una carga positiva. Los ionóforos de protones se estudiaron en principio como moléculas clasificadas funcionalmente como «desacopladores»: inhiben la fosforilación, pero estimulan la respiración. Este modo de acción se explica por su capacidad de disipar el gradiente de protones. Los ionóforos de protones son ácidos débiles que captan o ceden fácilmente un protón *(color)* y pueden difundir por la membrana.

El transferidor al que ceden el par de electrones las ferroproteínas sulfuradas es una pequeña molécula denominada ubiquinona, o coenzima Q. Frecuentemente se abrevia por Q. Mitchell da el nombre de ciclo Q a su mecanismo propuesto. La ubiquinona es una molécula cíclica que contiene dos átomos de óxigeno y posee tres estados de oxidación posibles. En el estado de máxima oxidación, o forma quinona, ambos oxígenos están unidos al anillo por dobles enlaces. Uniendo un átomo de hidrógeno a uno de los oxígenos se crea la forma semiquinona, o QH*. En la forma más reducida, llamada hidroquinona o QH_2, ambos oxígenos están unidos a átomos de hidrógeno.

En el ciclo Q propuesto, dos moléculas de ubiquinona captan respectivamente un electrón procedente de las ferroproteínas sulfuradas y toman protones del medio interno para formar dos moléculas de semiquinona, QH*. Dos electrones más son suministrados entonces por moléculas de otra proteína de la cadena respiratoria, el citocromo *b*. (La procedencia de éstos se discutirá más adelante.) Al captar dos protones más del medio intramitocondrial se forman dos moléculas de QH_2, la forma más reducida.

La ubiquinona es soluble en la matriz lipídica de la membrana, por lo que puede que sea móvil. A diferencia de los demás transferidores, que presumiblemente per-

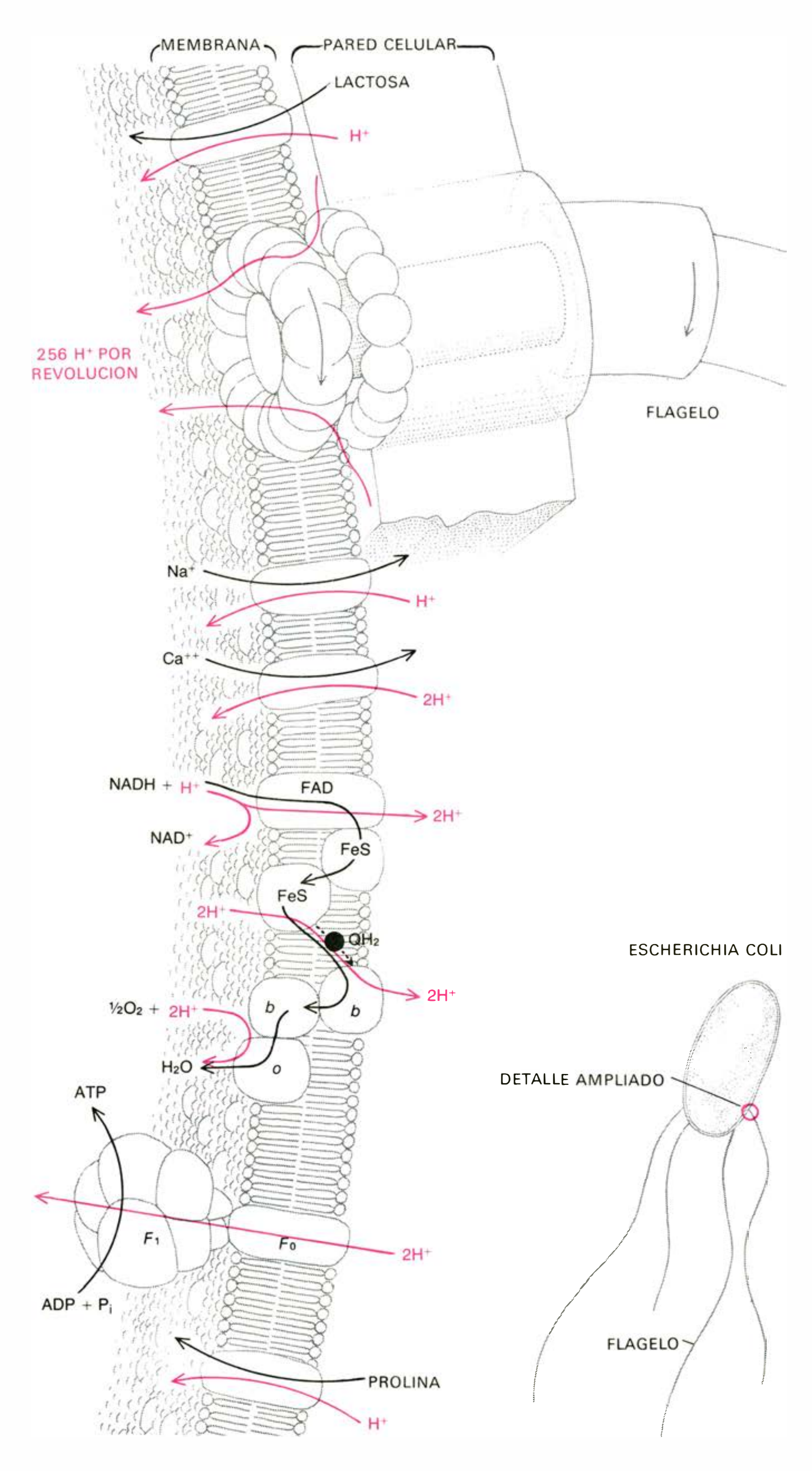


EN LA MEMBRANA DE LAS MITOCONDRIAS se encuentran incluidos los enzimas y otros componentes de la cadena respiratoria. La disposición de las moléculas no se conoce todavía con seguridad; el modelo presentado aquí es, en cierto modo, hipotético. La respiración comienza en el NADH, que cede dos electrones y un protón al flavín-mononucleótido (FMN); otro protón es captado del medio interno, de manera que la forma reducida de la molécula ($FMNH_2$) tiene dos átomos de hidrógeno completos. Los protones son expulsados y los electrones retornan a la superficie interna de la membrana a través de una ferroproteína sulfurada. Una vez aquí, los dos electrones son cedidos a dos moléculas de ubiquinona (Q), cada una de las cuales adquiere un protón y da lugar a la forma semiquinona (QH*). A diferencia de los otros componentes de la cadena respiratoria, las quinonas migran probablemente como moléculas completas a través de la membrana *(líneas a trazos)*. La semiquinona capta dos electrones más, del citocromo *b*, y con dos protones procedentes del interior de la mitocondria da lugar a la forma hidroquinona (QH_2), que es el estado más reducido. Cada hidroquinona cede uno al citocromo c_1, y libera al exterior el protón correspondiente. Los dos electrones restantes son devueltos al ciclo por medio del citocromo *b*, y se liberan los dos protones que restaban para completar el número de seis. Por último, los dos electrones cedidos al citocromo c_1 atraviesan los citocromos *c*, *a* y a_3 hasta alcanzar el oxígeno, que de esta forma es reducido a agua. El complejo F_1-F_0 completa el circuito de protones. En él, por cada dos protones introducidos, se sintetiza una molécula de ATP. El gradiente de protones también da energía a otros procesos, como por ejemplo la reducción del $NADP^+$ por el NADH, el transporte de iones de calcio (Ca^{++}) y sodio (Na^+), y el intercambio de ADP por ATP. Las ilustraciones de esta página se basan en dibujos hechos por Maija V. Hinkle.

manecen fijos y no sirven más que para conducir los átomos de hidrógeno o los electrones, la ubiquinona quizá migre como molécula de una a otra parte de la membrana. Apoya esta hipótesis la observación de que en la membrana hay aproximadamente diez veces más moléculas de ubiquinona que moléculas de otros componentes de la cadena respiratoria. En el ciclo Q propuesto, las dos moléculas de hidroquinona atraviesan la membrana desde la superficie interior, de donde toman un total de cuatro átomos de hidrógeno, hasta la exterior. Aquí, cada molécula de QH_2 cede un electrón a la siguiente proteína de la cadena respiratoria, el citocromo c_1, y libera un protón fuera de la mitocondria. En consecuencia, el número de protones transportados hasta el momento es de cuatro.

Todas las moléculas de ubiquinona se encuentran ahora en el estado de semiquinona, o QH*. Completan el ciclo retornando a la forma de máxima oxidación. Cada una cede su protón restante al medio externo y transfiere el electrón asociado al citocromo *b*. Los seis protones han sido ya trasladados a través de la membrana y expulsados de la mitocondria, pero resta aún por determinar el destino de cuatro electrones: dos fueron donados al citocromo c_1 y dos al citocromo *b*. Los citocromos son proteínas que contienen un grupo hemo: una estructura con un gran anillo en el que hay un átomo de hierro central que puede captar y ceder fácilmente un electrón. Los dos electrones son devueltos al ciclo Q a través de la membrana y finalmente reducen dos moléculas más de QH* a QH_2. Este círculo cerrado es el que proporciona los electrones necesarios para la reducción de las semiquinonas mencionada anteriormente.

Los dos electrones depositados en las moléculas de citocromo c_1 prosiguen por toda la cadena respiratoria, hasta su final. Son transferidos del citocromo c_1, incluido en la superficie exterior de la membrana, al citocromo *c* situado sobre la superficie ex-



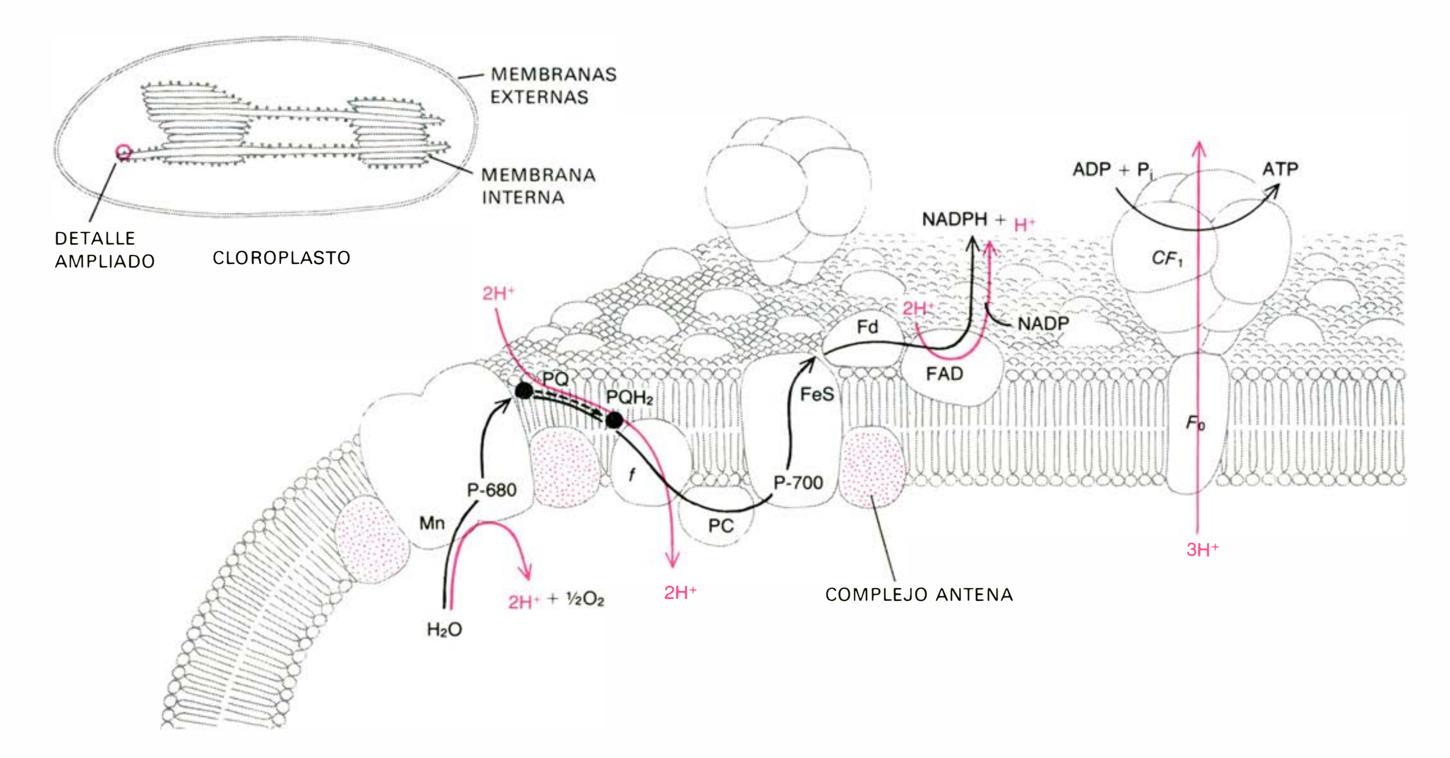
LA MEMBRANA DE LA BACTERIA *E. coli* contiene un sistema de proteínas respiratorias similares al sistema mitocondrial. Los electrones y los protones son cedidos al flavín-adenín-dinucleótido ($FADH_2$), que lleva al exterior los dos protones; los electrones son devueltos a la superficie interna por medio de ferroproteínas sulfuradas. Luego, los electrones, junto con dos protones procedentes del medio interno, reducen una sola molécula de ubiquinona a hidroquinona (QH_2), que difunde a través de la membrana y libera los protones en el exterior. Por último, los electrones prosiguen entre los citocromos *b* y *o* hasta el agua. Se transfieren, por tanto, sólo cuatro protones por cada par de electrones, en lugar de seis. Al igual que en las mitocondrias, la fosforilación requiere dos protones para cada molécula de ATP. El transporte de sodio y calcio, así como la captación de determinados azúcares, como la lactosa, y aminoácidos, como la prolina, van acoplados al movimiento de los protones a favor del gradiente. La rotación del flagelo toma energía de la entrada de protones. En la base del flagelo hay un anillo de 16 proteínas, adyacente a un anillo de la pared celular.

terior. Seguidamente, los dos electrones pasan al citocromo *a* y, atravesando la membrana por última vez, llegan al citocromo a_3. Por último, el citocromo a_3 es oxidado por oxígeno molecular. Los dos electrones son cedidos a un átomo de oxígeno y dos protones son captados del medio interno de la mitocondria con lo que se forma una molécula de agua.

Durante esta larga serie de reacciones de oxidación-reducción, el par de electrones atraviesa tres veces la membrana en ambos sentidos y extrae dos protones en cada una de las salidas. El primer paso lo hace por medio de la forma reducida del flavín-mononucleótido, $FMNH_2$. Después de retornar a través de las ferroproteínas sulfuradas, los electrones vuelven a salir con la hidroquinona. El segundo viaje de vuelta se efectúa por medio del citocromo *b*, los electrones reducen otro par de moléculas de ubiquinona y migran por tercera y última vez a la superficie exterior de la membrana. Finalmente vuelven al interior de la mitocondria a través del sistema de citocromos (c_1, c, a y a_3), y son suministrados al oxígeno. El flujo de electrones en la mitocondria supone una intensidad de corriente eléctrica. La diferencia de potencial entre el NADH y el oxígeno es de alrededor de 1,2 voltios, y la intensidad total en las mitocondrias de una persona en reposo es de aproximadamente 100 amperios, con lo que se generan 120 vatios de potencia.

Hay que hacer notar que esta reconstrucción de las proteínas de la membrana mitocondrial y de sus interacciones es hipotética. Sólo de unas pocas moléculas se tiene cierta evidencia que apoye su posición en la membrana. El FMN debe estar expuesto a la superficie interior, ya que sólo allí reacciona con el NADH. Por el contrario, el citocromo *c* debe ser exterior, ya que puede desprenderse lavando las mitocondrias con soluciones salinas. Se ha demostrado que dos agrupaciones de moléculas de citocromo, una compuesta de *b* y *c* y la otra de *a* y a_3, cruzan la membrana, aunque de momento no se ha establecido la posición exacta de cada citocromo. De todas formas, también otras disposiciones de las mismas moléculas podrían conducir al mismo resultado: que la transferencia de dos electrones del NADH al oxígeno vaya sincronizada con la salida de seis protones del interior al exterior de la mitocondria.

La cadena respiratoria de la bacteria *Escherichia coli* presenta varios componentes análogos a los encontrados en las membranas mitocondriales, y funciona de acuerdo con los mismos principios. Hay, sin embargo, una diferencia fundamental: por cada par de electrones que atraviesan la cadena respiratoria, sólo salen de la célula bacteriana cuatro protones (en lugar



LA MEMBRANA DEL CLOROPLASTO posee un sistema de pigmentos y otras moléculas dependientes de la luz, que transfieren protones al interior. Deben ser absorbidos dos fotones, o cuantos de luz, por cada electrón transferido desde el agua hasta el aceptor final de electrones, el NADP. Por cada par de electrones que atraviesa la cadena de moléculas transferidoras, tres protones son extraídos del cloroplasto y cuatro aparecen en el interior. El primer fotón es absorbido por una agrupación de moléculas de clorofila, el complejo antena, asociado a una clorofila especializada, que se designa P-680. Dos electrones procedentes de moléculas P-680 atraviesan la membrana y son reemplazados por electrones extraídos a una molécula de agua. Los electrones procedentes del P-680, junto con dos protones del exterior de la membrana, reducen la plastoquinona (PQ) a PQH_2. Los protones son liberados en el interior al ser transferidos los electrones al citocromo *f*. Seguidamente atraviesan la plastocianina (PC) hacia un segundo fotosistema, el P-700. Con la absorción de un fotón adicional, los electrones completan su recorrido pasando por una ferroproteína sulfurada (FeS), la ferredoxina (Fd) y el $FADH_2$ hasta el $NADP^+$. La síntesis de ATP en el complejo CF_1-F_0 es el único proceso que requiere un gradiente de protones; aparentemente tres protones cruzan la membrana por cada molécula de ATP que se forma.

de seis). El número de protones transferidos varía también según las condiciones en que hayan crecido las bacterias.

No sorprenden las similitudes entre bacterias y mitocondrias. Se considera que tanto las mitocondrias como los cloroplastos han evolucionado a partir de bacterias que penetraron en la célula como parásitos o simbiontes en los orígenes de las células nucleadas, y sólo posteriormente llegaron a ser orgánulos cautivos. Mitocondrias y cloroplastos poseen material genético propio, que, en su forma, se asemeja al bacteriano. Muchas proteínas de las mitocondrias y los cloroplastos se parecen más a proteínas bacterianas que a proteínas celulares de función equivalente.

En *E. coli* la cadena respiratoria es aparentemente más simple que la de las mitocondrias. Los electrones son suministrados a las proteínas respiratorias de la membrana celular por el NADH y son cedidos a un grupo flavínico análogo al FMN; el grupo es el flavín-adenín-dinucleótido, o FAD. Como el FMN, el FAD está asociado a una proteína y transporta dos átomos de hidrógeno a la superficie exterior de la membrana, donde son liberados los protones. Los electrones son transferidos a ferroproteínas sulfuradas que los devuelven a través de la membrana. Como en la respiración mitocondrial, el segundo transferidor de hidrógeno es la ubiquinona, aunque de momento no hay pruebas de un ciclo Q en *E. coli*. Los dos electrones son cedidos a una única molécula de quinona, que capta dos protones del interior de la célula formando la hidroquinona. Tras migrar a través de la membrana, la hidroquinona libera los dos protones en el exterior y transfiere los dos electrones a dos moléculas de citocromo *b*. Finalmente, los electrones retornan a través de la membrana por medio del citocromo *b* y del citocromo *o*, que es oxidado por oxígeno molecular. En resumen, el par de electrones cruza la membrana dos veces en cada dirección y transporta al exterior cuatro protones. En otras bacterias la cadena respiratoria toma formas distintas y en algunas es similar a la cadena mitocondrial. Debe destacarse que, además del NADH, pueden oxidarse otros compuestos en las mitocondrias y las bacterias.

En los cloroplastos, la cadena de transferencia electrónica tiene varias moléculas transferidoras similares a las de las cadenas respiratorias de mitocondrias y bacterias, pero el flujo de electrones se efectúa en el sentido contrario. La cadena comienza en el agua y finaliza en una forma fosforilada del NAD^+: el fosfato de nicotinamida-adenina-dinucleótido, o $NADP^+$. Más destacable aún es que el sentido de transporte de los protones es el contrario del que se da en las mitocondrias. Los cloroplastos tienen un sistema triple de membranas: las dos exteriores definen el orgánulo, y la interior, que está muy replegada, contiene el aparato de la fosforilación fotosintética. En los cloroplastos los protones son transportados al interior a través de la membrana interna y son acumulados dentro. Para que los electrones puedan fluir del agua al $NADP^+$ ha de elevarse la energía libre del estado inicial. La energía requerida la proporciona la absorción de luz por pigmentos lumínicos dispuestos en dos puntos de la cadena de transferencia de electrones. Cada fotón absorbido conduce un electrón a través de la membrana interna.

El primer fotón es absorbido por una agrupación de moléculas que incluye varios cientos de moléculas de clorofila y otros pigmentos, que están todos ligados a proteínas. Esta densa colección de substancias que absorben luz, denominado el complejo antena, está incluida en la superficie interna de la membrana. La energía de excitación se comunica rápidamente del complejo antena a una molécula de clorofila especializada, llamada P-680. («P» sustituye a «pig-

mento», y el número hace referencia a la longitud de onda de la luz que absorbe la molécula, expresada en nanómetros.) Las clorofilas tienen una estructura anular parecida a la del grupo hemo, salvo que la cavidad central la ocupa un átomo de magnesio en vez de otro de hierro. En P-680 la energía de la luz altera la distribución de los electrones en el anillo y permite que un electrón pueda ser transferido.

El electrón del P-680 es transferido aparentemente a un aceptor de electrones, situado en la superficie externa de la membrana, inmediatamente después de ser absorbido el fotón. El proceso se repite con la absorción de otro fotón. Las dos moléculas oxidadas de P-680 son reducidas por dos electrones, procedentes del agua, con la ayuda de un enzima que contiene manganeso. El átomo de oxígeno procedente de la molécula de agua se libera y sale fuera del cloroplasto; los dos protones pasan a la solución en el interior de la membrana interna. Los dos electrones que atraviesan la membrana procedentes del P-680 son captados en la superficie externa por un transferidor de hidrógeno, de estructura análoga a la de la ubiquinona, pero denominado plastoquinona, o PQ. Los dos electrones, con dos protones extraídos de la solución exterior reducen la PQ a PQH_2, que retorna a través de la membrana a la superficie interna. Una vez allí la PQH_2 libera los protones al interior y transfiere los electrones al citocromo *f*, una proteína del mismo tipo que el citocromo *c*.

El siguiente transferidor de electrones de la cadena es la proteína plastocianina, que contiene cobre, y que lleva los electrones al segundo sistema fotoquímicamente activo. El componente más importante de

NADH — FLAVINA — UBIQUINONA

FORMA MAS REDUCIDA

FORMA OXIDADA

(FORMA HIDROQUINONA)

(FORMA SEMIQUINONA)

(FORMA QUINONA)

CENTRO FERRO-SULFURADO

● = Fe

GRUPO HEMO *a*

CLOROFILA *a*

LAS MOLECULAS TRANSFERIDORAS, tanto de la cadena respiratoria como de la fotosintética, están especializadas en el transporte de electrones o átomos de hidrógeno. Se muestran aquí los centros activos de algunos transferidores; la *R* quiere indicar el resto de moléculas. El átomo (o átomos) de hidrógeno cedido a las formas oxidadas de los transferidores para dar lugar a las formas reducidas figura en color. El NAD^+ toma dos electrones y un protón para dar NADH. Un electrón neutraliza la carga del átomo de nitrógeno del anillo, y el otro se combina con un protón para hidrogenar un átomo de carbono. Una molécula muy similar, el $NADP^+$, hace de transferidor en los cloroplastos. Las flavinas FAD y FMN forman las partes activas de las flavoproteínas. Dos átomos de hidrógeno son cedidos al FAD y al FMN para dar lugar respectivamente a $FADH_2$ y $FMNH_2$. Las ferroproteínas sulfuradas son transferidores de electrones, más que de átomos de hidrógeno; el centro activo de una de ellas es una estructura con forma de jaula en la que la carga de los electrones se distribuye entre los átomos de hierro y de azufre. Algunos de los átomos de azufre están en forma de sulfuro (S^{--}), y el resto de ellos proceden del aminoácido cisteína de la proteína (*R*). La ubiquinona es un transportador de hidrógeno en las membranas de mitocondrias y bacterias. En su estado de máxima reducción y protonación los átomos de oxígeno del anillo bencénico están unidos a átomos de hidrógeno. En el estado de máxima oxidación los átomos de oxígeno están unidos al anillo de carbono mediante dobles enlaces. La ubiquinona puede presentarse también en una forma semirreducida, denominada semiquinona, en la que un oxígeno se une a un átomo de hidrógeno. La semiquinona es un radical libre en que el electrón desapareado se desplaza por resonancia en el anillo de carbono. La plastoquinona, un transferidor de electrones de los cloroplastos, es muy similar a la ubiquinona. La clorofila y el grupo hemo de los citocromos también son de estructura similar. En los citocromos, el electrón suele llevarlo el átomo de hierro (Fe) que ocupa el centro del gran anillo hemo; la carga positiva de la forma oxidada de la molécula de clorofila se distribuye, con probabilidad, por todo el anillo hemo.

este sistema es otra molécula especializada de clorofila, designada por P-700. Al excitarse el P-700 por absorción de luz en el complejo antena, los electrones son trasladados de nuevo a la superficie externa de la membrana, donde son aceptados por una ferroproteína sulfurada. El camino prosigue por la superficie externa yendo de esta ferroproteína sulfurada a otra distinta, la ferredoxina. De aquí, los electrones van a un FAD con protones. Este toma dos protones del medio externo formando $FADH_2$. Finalmente, el $NADP^+$ toma los electrones y un protón del $FADH_2$, y forma NADPH. Cuando se ha completado el paso de dos electrones por la cadena, tres protones han quedado por fuera de la membrana y cuatro han aparecido en el interior.

Este modelo de la membrana interna del cloroplasto, al igual que los modelos de las membranas mitocondrial y bacteriana discutidos anteriormente, es una reconstrucción hipotética basada en pruebas sólo fragmentarias. Los estudios sobre la estructura de las membranas han mostrado que la ferredoxina se encuentra en la parte externa y que la plastocianina funcional está en la superficie interna. Las demás características del modelo son más especulativas. H. T. Witt y sus colegas del Instituto Max Volmer de Berlín han probado que se crea un potencial eléctrico a través de la membrana interna del cloroplasto en menos de 10^{-8} segundos después de una iluminación. La velocidad de la reacción es una prueba vigorosa de que los electrones liberados por las clorofilas especializadas son transferidos a través de la membrana; cualquier otro método menos directo de crear un potencial resultaría demasiado lento.

El gradiente de protones, establecido por medio del transporte fotoquímico u oxidativo de electrones, representa un almacenamiento de energía libre, algo así como el agua bombeada hasta un depósito elevado en contra de un gradiente gravitacional. La energía puede recuperarse permitiendo el flujo inverso de los protones a través de la membrana, por entre un «molino» adecuado. El trabajo principal que efectúa el flujo de protones a través del gradiente es la fosforilación del ADP a ATP, pero el gradiente también potencia otros procesos, como el del transporte activo de ciertos iones.

La energía se almacena en el gradiente de protones de dos formas, dicho en otras palabras, el gradiente tiene dos componentes. Un componente es la diferencia de concentración o actividad química de los protones a ambos lados de la membrana. Los protones tienden a difundir de una región con alta concentración a otra región con baja concentración cuando se establece una vía por entre la membrana. La concentración de protones se mide en unidades de *p*H, definido como el logaritmo, cambiado de signo, de la concentración de iones de hidrógeno. La energía del gradiente de concentración se determina por la diferencia de *p*H a través de la membrana, y es independiente de la magnitud absoluta del *p*H.

La carga eléctrica transportada por los protones da lugar al segundo componente de la energía del gradiente. El movimiento neto de las cargas a través de la membrana crea una diferencia de potencial eléctrico y todas las partículas cargadas quedan afectadas por el campo electrostático resultante. La energía total del gradiente de protones es la suma del componente de concentración (u osmótico) y el componente eléctrico.

Debido a la diferencia de concentración y de potencial eléctrico, un protón expulsado de una mitocondria experimenta una fuerza que tiende a introducirlo de nuevo. El movimiento del protón en respuesta a esta fuerza puede utilizarse para efectuar un trabajo, como, por ejemplo, el de la fosforilación.

Los enzimas que ajustan la difusión de retorno de protones por la membrana con la síntesis de ATP parecen ser notablemente similares en mitocondrias, bacterias y cloroplastos. Destacan en las microfotografías electrónicas como cuerpos globulares que sobresalen de la superficie de la membrana. El prominente «botón» designado por F_1 fue aislado, en 1960, por Efraim Racker, Maynard E. Pullman y Harvey Penefsky, del Public Health Research Institute of the City of New York. Es una proteína soluble compuesta de cinco tipos de subunidades, algunas de las cuales se presentan repetidamente. F_1 se engancha a la membrana por medio de otro conjunto de proteínas, designado por F_0, incluido en la membrana y que parece que la atraviesa totalmente.

F_1 puede desprenderse fácilmente de la membrana, pero F_0 sólo queda libre cuando la membrana es destruida mediante detergentes. El sistema completo de enzimas se denomina el complejo F_1-F_0.

En las mitocondrias y bacterias el complejo F_1-F_0 se orienta de forma que las prominencias sobresalgan hacia la matriz interior desde la superficie interna de la membrana. Está comprobado que por cada dos protones que penetran por entre el complejo, una molécula de ADP se combina con fosfato inorgánico formando ATP. Es importante hacer notar que la reacción es reversible: en circunstancias apropiadas el complejo F_1-F_0 puede disociar moléculas de ATP y utilizar la energía desprendida para bombear protones hacia el exterior de la mitocondria o célula bacteriana. Como todos los enzimas, el complejo F_1-F_0 controla la velocidad de la reacción, pero el sentido de ella queda determinado por el balance de energía libre.

En los cloroplastos, el sistema enzimático equivalente se denomina complejo CF_1-F_0, y está orientado en sentido contrario, es decir, la prominencia sobresale de la superficie externa de la membrana. La orientación es exactamente la que cabía esperar, ya que el sentido del gradiente de protones en los cloroplastos es también contrario al de mitocondrias y bacterias. Los protones fluyen hacia fuera por entre el complejo CF_1-F_0 y el ATP se forma en el espacio que queda entre las membranas interna y externa del cloroplasto. Otra diferencia es más difícil de justificar: el complejo CF_1-F_0 del cloroplasto requiere aparentemente tres protones por cada ATP sintetizado, en lugar de dos.

No se conoce el mecanismo de síntesis de ATP en el locus activo del complejo F_1-F_0. Se han formulado varias hipótesis, pero los estudios del enzima no han proporcionado razones suficientes como para decidir entre ellas.

El locus activo está aparentemente asociado a la parte F_1 del complejo, las prominencias observables en las microfotografías electrónicas. Aun estando aislados y en solución, los componentes del F_1 catalizan la disociación del ATP que da lugar al ADP y fosfato, reacción inversa a la de fosforilación originada por un gradiente de protones. La reacción en el sentido directo no puede naturalmente observarse con moléculas aisladas de F_1, ya que en solución no puede establecerse un gradiente de protones. De los cinco tipos de subunidades proteicas de F_1, se requieren sólo dos de las de mayor tamaño, para que haya actividad disociadora del ATP. La menor de las subunidades es un inhibidor de la actividad catalítica, que presumiblemente sirve de válvula de seguridad: evita que la reacción prosiga cuando el gradiente de protones es demasiado pequeño, ya que en estas condiciones la reacción discurriría a la inversa, consumiendo ATP.

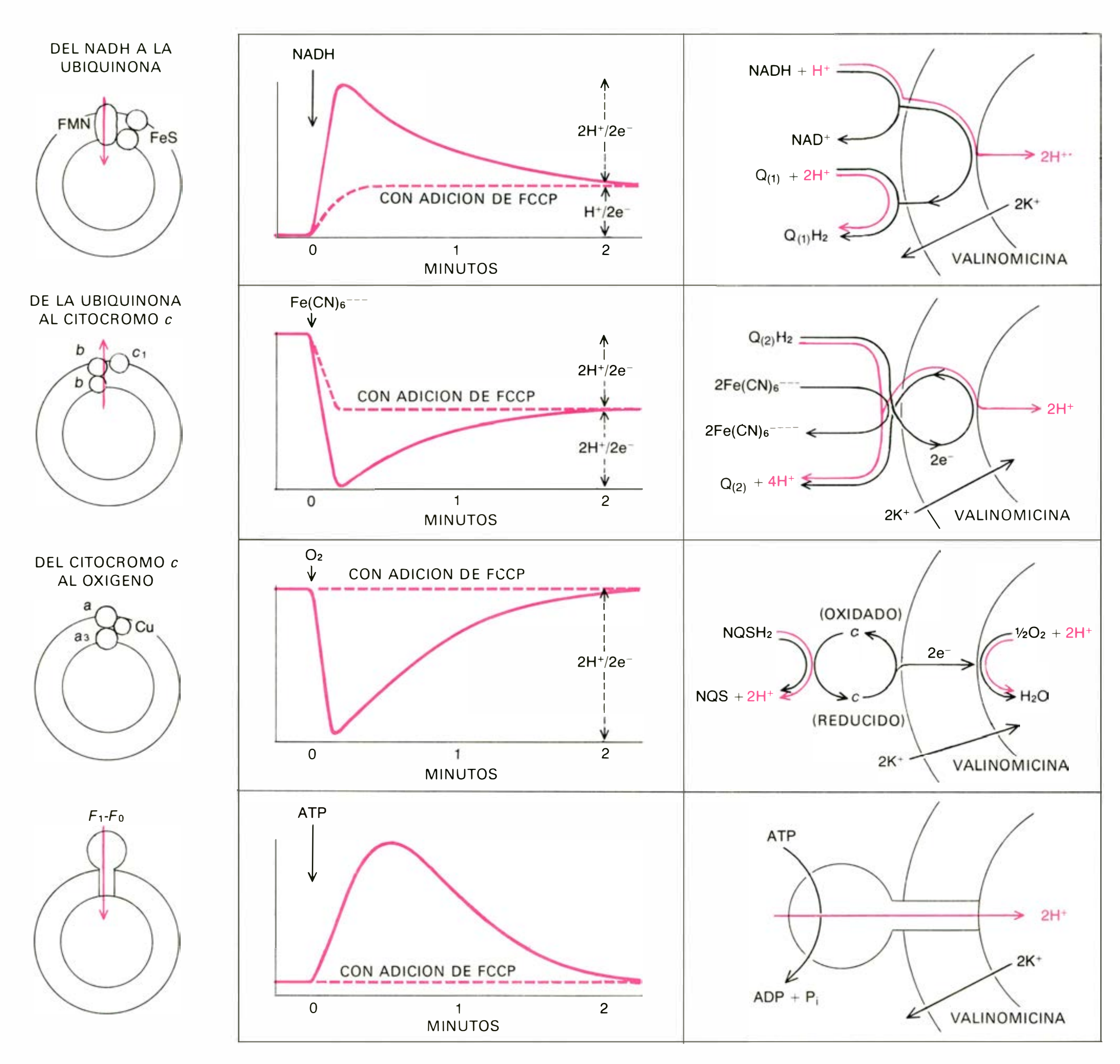
La parte F_0 del complejo, que permanece en la membrana al extraer F_1, no muestra actividad catalítica con el ATP. Por otro lado, la extracción de F_1 hace a la membrana altamente permeable a los protones y estos huecos pueden obturarse reponiendo F_1, o por medio de ciertos inhibidores. Estas observaciones han llevado a la interpretación de que el F_0 actúa como un canal por el que los protones atraviesan la membrana hasta F_1.

Uno de los hipotéticos mecanismos catalíticos fue propuesto por Mitchell. Según este esquema, un grupo fosfato se une al enzima en el locus activo, dentro de la parte F_1 del complejo, pero cerca del final del canal de protones de F_0. Dos protones impulsados por el canal mediante el po-

tencial de membrana y el gradiente de *p*H atacan a uno de los oxígenos del fosfato provocando su pérdida para formar una molécula de agua. El enlace fosfórico libre creado de esta manera transforma el grupo fosfato en una especie altamente reaccionante que puede enlazarse directamente al ADP.

En las principales hipótesis alternativas los protones desempeñan un papel menos directo. Estas hipótesis indirectas adoptan entonces varios mecanismos posibles. La idea común a todas ellas es que el paso de protones por la parte F_1 del complejo podría cambiar la conformación proteica. Por ejemplo, la unión de un protón a un grupo de la proteína podría causar su desplazamiento debido a interacciones eléctricas con otras zonas cargadas. Paul D. Boyer, de la Universidad de California en Los Angeles, ha propuesto que los cambios inducidos por protones en la conformación proteica en las proximidades del locus activo podrían dar lugar a la síntesis de ATP. En un mecanismo de este tipo la separación de la molécula ya completa de ATP y el enzima es el paso crucial en donde se requiere la energía. Aunque el ADP libre no puede combinarse con el fosfato sin un aporte de energía, la reacción podría tener lugar espontáneamente si ambas moléculas estuvieran ligadas a una proteína. El ATP resultante permanecería unido al enzima y sólo podría separarse con la adición de energía. Esta energía podría aportarse mediante un cambio conformacional, basado a su vez en la transferencia de protones.

En 1971, I. J. Ryrie y André T. Jagendorf, de la Universidad de Cornell, obser-



PARTES DEL SISTEMA DE ENZIMAS de las membranas mitocondriales pueden aislarse y reconstruirse en vesículas de membrana artificiales; al suministrar los substratos adecuados, se observa que los segmentos transportan protones a través de la membrana. El primer segmento consta de los componentes de la cadena respiratoria que van desde el NADH hasta la ubiquinona, incluyendo el FMN y las ferroproteínas sulfuradas. El segundo segmento incluye las moléculas situadas entre la ubiquinona hasta el citocromo *c*. El tercer segmento completa la cadena respiratoria desde el citocromo *c* hasta el oxígeno. El cuarto sistema aislado es el complejo F_1-F_0, que opera en sentido contrario, es decir, disociando ATP para formar el gradiente. Los sistemas respiratorios pueden dirigirse añadiendo un dador y un aceptor de electrones, que pueden ser substratos naturales, como el NADH o el oxígeno, o artificiales, como el ferrocianuro ($Fe(CN^{---})6$), moléculas análogas a la ubiquinona (Q_1 o Q_2) o el reductor naftoquinol sulfonato ($NQSH_2$). El transporte de protones se mide por medio del *p*H del medio externo. En algunos casos, la reacción da lugar a un cambio permanente en el *p*H, que es independiente del gradiente, pero al no estar afectado por el ionóforo de protones FCCP, puede medirse por separado. Si se corrige esta derivación, la cadena respiratoria transporta dos protones por par de electrones.

varon cambios en la conformación del CF_1 unido a cloroplastos cuando éstos eran iluminados, o cuando el *p*H del medio en el que estaban suspendidos era cambiado rápidamente de ácido a básico, creando un gradiente de *p*H momentáneo. R. P. Magnusson, J. Fagan y uno de nosotros (McCarty) mostraron que la reactividad de un grupo químico en una de las subunidades del CF_1 aumentaba al iluminar los cloroplastos. Penefsky y otros han demostrado que la proteína F_1 de las mitocondrias sufre también cambios estructurales al ser expuesta a un gradiente electroquímico de protones. En 1973, E. C. Slater y sus colegas, de la Universidad de Amsterdam, hallaron que CF_1 y F_1 poseen ADP y ATP firmemente unidos a ellos. En los cloroplastos puede intercambiarse el ADP unido al CF_1 por nucleótidos del medio, al formarse un gradiente de protones. El ADP enlazado puede fosforilarse, pero sólo en una baja proporción. Probablemente no debe estar unido a un locus catalítico del enzima, pero debe regular la estructura o el funcionamiento de las moléculas de F_1.

Estos hallazgos pueden interpretarse como la prueba de un mecanismo de fosforilación, en el que los cambios conformacionales desempeñan un papel esencial. No obstante, la hipótesis directa de Mitchell no puede descartarse. Los cambios conformacionales raramente escasean y pueden considerarse universales en los enzimas. Los cambios en F_1 podrían estar asociados a un mecanismo directo o ser propios de la actividad catalítica.

El movimiento de moléculas o iones a través de una membrana puede producirse espontáneamente sólo en el sentido de que tiende a reducir un gradiente de concentración o un gradiente de potencial eléctrico. En las células y en muchos orgánulos subcelulares muchas substancias deben ser asimiladas o expulsadas en contra de un gradiente, y por ello la mayoría de los transportes requieren energía. El movimiento de algunos iones y moléculas está dirigido directamente por el gradiente de protones.

Un ejemplo de transporte ligado a protones es el flujo de iones sodio (Na^+) hacia el exterior de la mitocondria, a cambio de iones de hidrógeno. En este sistema un ion sodio sale de la mitocondria por cada protón que atraviesa la membrana en sentido contrario. Se supone que el intercambio está mediado por una proteína de la membrana.

El transporte de sodio está dirigido por sólo un componente del gradiente electroquímico de protones, a saber, la diferencia de concentración de protones, o *p*H. El potencial de membrana no afecta, ya que, aun cuando acelera el paso de protones hacia el interior, retarda el movimiento de salida de los iones sodio, con lo que los dos efectos se anulan. La reacción alcanza el equilibrio cuando el gradiente de sodio es equivalente al gradiente de *p*H en concentración de protones. Un sistema de transporte equivalente potencia la captación de fosfato inorgánico, que penetra en la mitocondria en forma de ion cargado negativamente. El movimiento de los iones fosfato queda compensado por un contraflujo de iones hidroxilo (OH^-).

El otro componente del gradiente de protones, el potencial de membrana, puede también dirigir un transporte activo. La captación de iones calcio (Ca^{++}) por la mitocondria, por ejemplo, toma la energía del potencial de membrana y es independiente del gradiente de *p*H. El calcio entra probablemente sin contracorriente de algún otro ion, estando motivado su transporte sólo por fuerzas electrostáticas que lo atraen a la superficie interna de la membrana, cargada negativamente. El potencial de membrana proporciona también la fuerza que motiva el intercambio de moléculas de ADP y ATP, asegurando un abastecimiento constante de substrato para la fosforilación y exportando el producto resultante. El ADP lleva una carga de menos 3 y el ATP otra de menos 4, de forma que el intercambio es equivalente a la salida de una carga negativa o la entrada de otra positiva.

El gradiente de protones de los cloroplastos no está aparentemente acoplado a ninguna función distinta que no sea la de fosforilación, con la posible salvedad del transporte de iones magnesio. Por el contrario, en las bacterias una gran variedad de procesos metabólicos parecen derivar su energía del gradiente de protones. De hecho, Mitchell se encontraba estudiando el transporte a través de la membrana bacteriana cuando propuso, en 1961, la teoría quimiosmótica.

El mecanismo de transporte de sodio en *E. coli* es aparentemente similar al mitocondrial. En cambio, el transporte de calcio tiene lugar en el sentido opuesto, es decir, *E. coli* excreta calcio. El transporte de calcio en las bacterias es probablemente un intercambio eléctricamente neutro de dos protones por cada ion de calcio, movimiento potenciado por el gradiente de *p*H.

En 1963, Mitchell propuso que la captación del azúcar lactosa por *E. coli* podría estar acoplada a una entrada de protones, aun cuando la lactosa más que un ion es una molécula eléctricamente neutra. Estudios posteriores han mostrado que el transporte de lactosa puede dirigirse por un gradiente de protones artificial y que, recíprocamente, la entrada de lactosa en células desnutridas va acompañada de una entrada de protones. El sistema de transporte parece ser simétrico: el azúcar puede concentrarse a ambos lados de la membrana, dependiendo de la polaridad del gradiente de protones. Sistemas similares para la captación de ciertos aminoácidos han sido descubiertos en *E. coli*.

Hay que resaltar que no son éstos los únicos sistemas de transporte activo de las bacterias; hay al menos dos más que no están directamente conectados al gradiente de protones. La glucosa y algunos otros azúcares son introducidos en las células bacterianas mediante un sistema en el que el fenómeno preliminar esencial es la fosforilación del azúcar. Otras moléculas atraviesan la membrana uniéndose primero a proteínas solubles en un sistema que toma de la disociación del ATP la energía necesaria para el transporte activo.

Una extraordinaria conexión mecánica dirigida por el gradiente de protones es el flagelo responsable de la motilidad de *E. coli* y otras bacterias. En la actualidad se cree que el flagelo se mueve girando como una hélice. El filamento a nivel de la base del flagelo atraviesa un cojinete en la pared celular rígida y finaliza en un rotor situado en la membrana celular. El rotor parece ser un anillo de 16 proteínas, y está yuxtapuesto a otro anillo proteico que quizás actúe de «estator» en la pared celular.

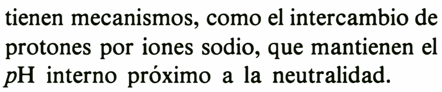
Howard C. Berg y sus colaboradores, de la Universidad de Colorado, han determinado recientemente que el rotor flagelar toma la energía del gradiente electroquímico de protones. El cálculo efectuado por Berg sobre la energía requerida por la rotación sugiere que deberían entrar 250 protones para potenciar una revolución, si el gradiente es el típico de *E. coli*. Los protones fluyen presumiblemente a través del anillo de proteínas que forma el rotor, generando una fuerza en contra del anillo adyacente en la pared celular. Una hipótesis obvia es que el paso de un protón a través de cada una de las 16 subunidades del anillo da un giro al flagelo de 1/16 de vuelta. Una revolución completa requeriría, por tanto, 16 × 16, es decir, 256 protones.

Al proponer Mitchell la teoría quimiosmótica se desconocía la traslocación de protones en las mitocondrias y cloroplastos. (Es más, él había sugerido primero que la respiración bombeaba protones al interior de la mitocondria en vez de hacia fuera.) La existencia de un sistema de transporte de protones acoplado a la respiración fue establecido por Mitchell y Moyle, que utilizaron una técnica, comparativamente más sencilla, para medir el número de protones transferidos por cada par de electrones. Se suspenden las mitocondrias en un medio sin oxígeno; entonces se verifica la concentración de iones hidrógeno tras inyectar una cantidad conocida de oxígeno a la solución. Se observa

que la concentración de protones aumenta inmediatamente y con gran velocidad; al agotarse el oxígeno, la concentración va retornando lentamente a un nivel normal, cuando los protones van de nuevo atravesando la membrana. Si se añade una substancia denominada ionóforo de protones, el gradiente se disipa inmediatamente: se sabe que estos ionóforos hacen la membrana permeable a los protones. En los cloroplastos se observa un comportamiento inverso. Tras un destello de luz, la concentración de protones en el medio externo decrece rápidamente y luego lentamente recupera el nivel original al salir los protones del cloroplasto. También aquí los ionóforos de protones colapsan el gradiente.

La medición de la magnitud del gradiente se complica por la necesidad de determinar la diferencia de *p*H y el potencial de membrana por separado. El gradiente de concentración de protones puede determinarse midiendo la entrada de un ácido débil en las mitocondrias o bacterias; en los cloroplastos se utiliza una base débil en lugar del ácido. El potencial de membrana queda de manifiesto por la captación de un ion (que no sea H^+ u OH^-) que atraviese la membrana libremente y alcance un gradiente de concentración de equilibrio.

Se ha hallado que, en las mitocondrias, el gradiente de protones consiste en una diferencia de *p*H de alrededor de 1,4 unidades de *p*H (ácido en el exterior), y un potencial de membrana de unos 140 milivoltios (positivo en el exterior). Las vesículas obtenidas a partir de membranas de *E. coli* tienen un gradiente de *p*H de dos unidades y un potencial de membrana de alrededor de 70 milivoltios, con la misma polaridad que en las mitocondrias. La manera en que el gradiente electroquímico se distribuye en sus dos componentes depende del tipo de iones presentes en la solución y del *p*H externo. En los cloroplastos, por supuesto, el gradiente se orienta en sentido contrario, y más aún, en el estado de equilibrio, casi todo el gradiente de los cloroplastos se manifiesta como una diferencia en la concentración de protones. En estas condiciones el potencial de membrana es insignificante, ya que la membrana es permeable a los iones cloruro, pero el gradiente de *p*H alcanza una magnitud de unas 3,5 unidades. En consecuencia, la matriz interior del cloroplasto puede alcanzar una fuerte acidez llegando hasta un *p*H de aproximadamente 4, y el cloroplasto acumula ácido clorhídrico en su interior. Debe señalarse de paso que los enzimas de la fijación del dióxido de carbono se encuentran entre las membranas interior y exterior y, por tanto, no están expuestos a este duro ambiente. Por el contrario el interior de las mitocondrias y bacterias es el lugar de muchas reacciones enzimáticas. Las mitocondrias y las bacterias

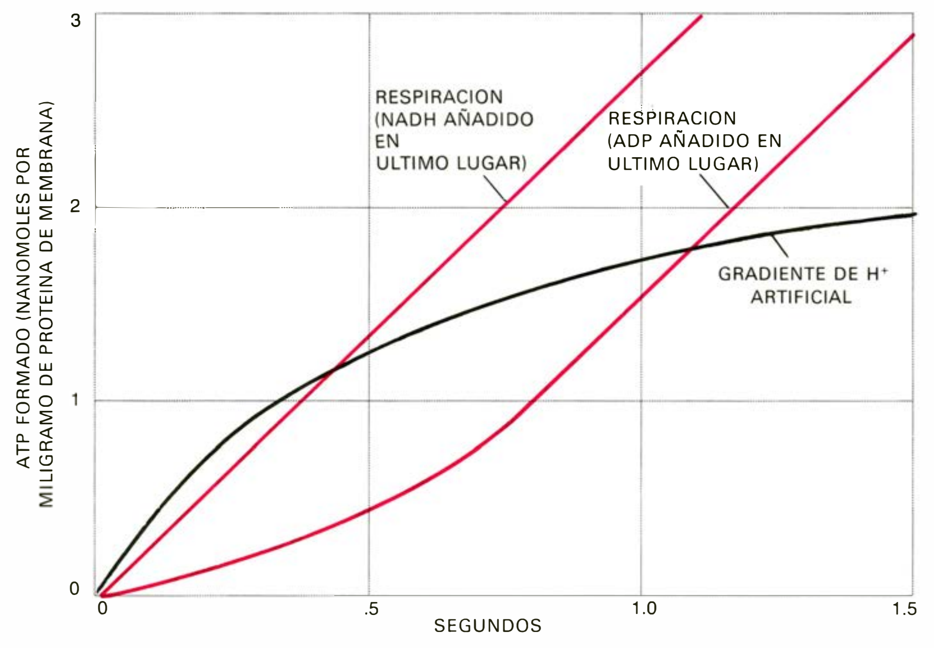


tienen mecanismos, como el intercambio de protones por iones sodio, que mantienen el *p*H interno próximo a la neutralidad.

El gradiente electroquímico completo, sin considerar su subdivisión en componentes de diferencia de *p*H y potencial de membrana, puede medirse en términos de una energía. Es la energía adquirida por un protón al ser transportado a través de la membrana en contra del gradiente, o la liberada por un protón al retornar por la membrana a favor del gradiente. En los cloroplastos el gradiente equivale a unas 4,8 kilocalorías por mol de protones; en las vesículas obtenidas a partir de membranas de *E. coli* es de 4,4 kilocalorías por mol. Las mitocondrias tienen el mayor de los gradientes totales, unas 5,3 kilocalorías por mol de protones.

La detección de un gradiente de protones generado por la respiración no es de ninguna manera una demostración de que la fosforilación esté dirigida por el gradiente. Los ionóforos de protones, no obstante, aportan una prueba que apoyan esta idea. Del dinitrofenol, primer ionóforo que se descubrió, se conocía su efecto sobre la oxidación fosforilativa mucho antes de comprenderse su mecanismo de acción. En 1948, William F. Loomis y Fritz A. Lipmann, de la Harvard Medical School, demostraron que el dinitrofenol inhibe la síntesis de ATP, pero estimula el transporte de electrones; dicho en otras palabras, desacopla la respiración de la fosforilación. En 1963, Mitchell propuso que el dinitrofenol actúa transfiriendo protones a través de las membranas, y demostró que la adición de dinitrofenol a bacterias equilibra un gradiente de *p*H a través de la membrana. Se ha probado desde entonces que la efectividad de diversos desacopladores está correlacionada con su capacidad de actuar como ionóforos. Una hipótesis alternativa es la de que los desacopladores actúan interfiriendo con el complejo F_1-F_0, pero el más potente de los desacopladores, una molécula designada SF6847, es completamente efectiva a una concentración de aproximadamente una molécula por cada cinco complejos F_1-F_0.

Para una desacoplación total deben ser disipados tanto el gradiente de *p*H como el potencial de membrana. Los ionóforos de protones pueden lograrlo, ya que transportan ellos mismos los protones a través de la membrana. Cada uno de los componentes del gradiente puede ser anulado individualmente por otros ionóforos. La valinomicina, un ionóforo que transfiere iones potasio (K^+), sólo disipa el potencial de membrana. En presencia de la valinomicina, una contracorriente de iones potasio neutraliza la diferencia de potencial eléctrico, pero si la respiración continúa se produce un au-



LA VELOCIDAD DE LA FOSFORILACION parece estar limitada por la magnitud del gradiente de protones. Se midió la velocidad de síntesis de ATP en vesículas obtenidas de membranas mitocondriales. Al aplicar un gradiente electroquímico de protones artificial, la velocidad inicial de síntesis era incluso mayor que la observada en condiciones naturales, pero la velocidad decrecía al ir disminuyendo el gradiente. Si se creaba un gradiente por respiración, se podía obtener una velocidad constante de síntesis, pero la velocidad inicial dependía de las condiciones experimentales. Si se suministraban todos los substratos necesarios para la respiración, salvo el ADP, al añadir éste comenzaba la fosforilación inmediatamente y a una velocidad elevada y constante. Si el substrato que faltaba era el NADH, se producía un retraso de alrededor de un segundo hasta alcanzar la máxima velocidad de síntesis. Presumiblemente era necesario este tiempo para que la cadena respiratoria formase un gradiente de protones. Estos hechos indican que un gradiente de protones puede suministrar la energía con la velocidad suficiente como para satisfacer los requerimientos de la fosforilación.

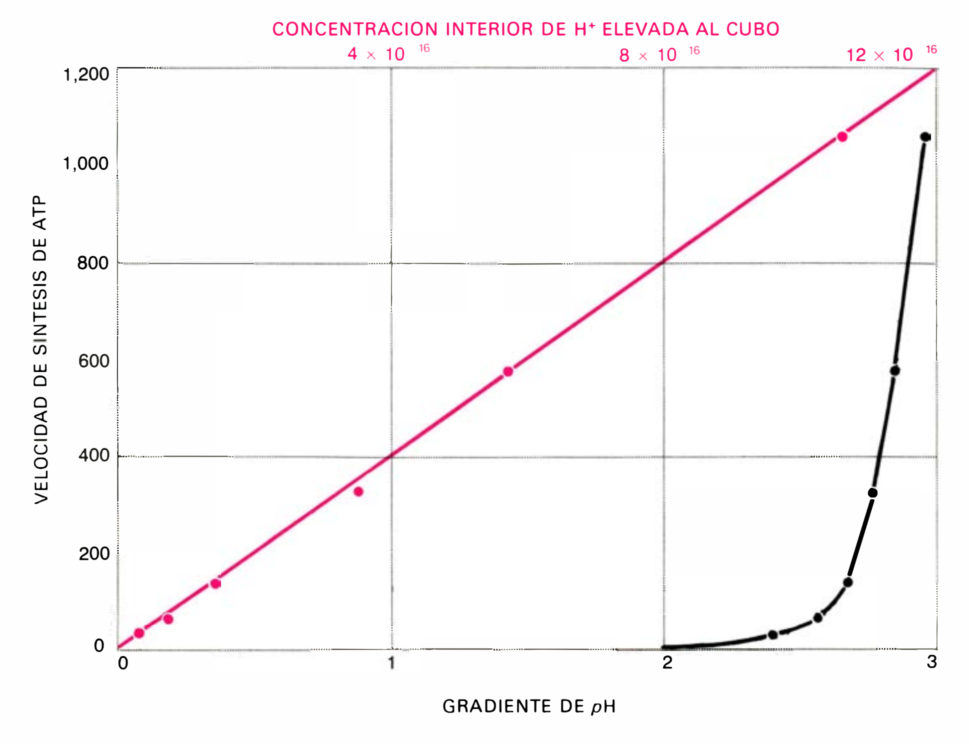
mento compensatorio del gradiente de *p*H. La nigericina, otro ionóforo, lleva a cabo un intercambio eléctricamente neutro de protones por iones sodio. Con ello elimina el gradiente de *p*H, de forma que una respiración continua aumenta el potencial de membrana. La fosforilación puede proseguir en presencia de cualquiera de los ionóforos. En presencia de ambos, no obstante, quedan colapsados los dos componentes del gradiente de protones, inhibiéndose fuertemente la fosforilación.

Cuando está presente un ionóforo de protones, la energía de la respiración se disipa en forma de calor. En las mitocondrias del tejido adiposo pardo este efecto se aprovecha para mantener la temperatura corporal. En estas mitocondrias una proteína de la membrana deshace el gradiente de protones; cabe el que los ácidos grasos actúen entonces como ionóforos.

Una demostración convincente del papel desempeñado por el gradiente de protones en la fosforilación la logró, en 1966, Jagendorf y Ernest G. Uribe, de la Universidad de Johns Hopkins. Demostraron que la creación de un gradiente artificial de *p*H puede dar lugar a la síntesis de ATP. Se suspendieron membranas internas de cloroplastos en una solución a *p*H 4 (moderadamente ácido), y el *p*H de la solución fue rápidamente elevado a 8 (moderadamente básico), de forma que los protones se vieron expulsados. Se formó entonces ATP en una cantidad correspondiente a 100 moléculas por cada complejo CF_1. Mitchell informó posteriormente de la síntesis de ATP en mitocondrias expuestas a una variación en sentido contrario, de medio básico a ácido. William S. Thayer, de Cornell, y uno de nosotros (Hinkle) logramos también la síntesis de ATP imponiendo gradientes artificiales a vesículas invertidas obtenidas a partir de membranas mitocondriales. La producción máxima de 10 ATP por cada F_1 se obtuvo al combinar un gradiente de *p*H con un potencial de membrana.

Si un gradiente electroquímico de protones ha de servir de mediador, acoplando la respiración a la fosforilación, debe ser energética y cinéticamente competente para lograrlo. Deben transferirse suficientes protones para alcanzar la cantidad de ATP sintetizada, los protones deben tener suficiente energía para suministrar la energía libre de la síntesis y han de corresponderse las tasas de transferencia y fosforilación.

Una suposición fundamental de los modelos de membrana presentados era que se transferían dos protones por cada «lugar de acoplamiento» o cada paso de un par de electrones a través de la membrana. De esta forma en la mitocondria, en que se supone que hay tres lugares de acoplamiento, deben transferirse seis protones por cada par de electrones que pasan del NADH al



LA CINETICA DE LA FOTOFOSFORILACION fue estudiada por medio de cloroplastos. Se determinó separadamente la velocidad de síntesis del ATP y la magnitud del gradiente de protones, pero bajo condiciones similares, en cloroplastos iluminados con diversas intensidades de luz. La relación derivada de estas mediciones *(curva y escala negras)* muestra que la velocidad de fosforilación depende fuertemente del gradiente de *p*H. En particular, no se fabrica ATP hasta que la diferencia de *p*H alcanza dos unidades de *p*H, presumiblemente porque los protones que se desplazan a favor de un gradiente menor no llevan suficiente energía como para desencadenar la reacción. Las mediciones de la velocidad quedan alineadas si se representan como función del cubo de la concentración interna de protones *(curva y escala en color)*. Esta relación de tercera magnitud apoya la prueba de que se requieren tres protones por molécula de ATP que se forma en los cloroplastos.

oxígeno. Se dice entonces que las mitocondrias tienen una proporción H^+/O, o una proporción $H^+/2e^-$, de 6. *E. coli* y los cloroplastos, en los que se supone que hay dos lugares de acoplamiento, tienen una proporción $H^+/2e^-$ de 4.

El estudio de tales proporciones en las reacciones químicas se denomina estequiometría. La estequiometría de la traslocación de protones puede determinarse experimentalmente en las mitocondrias y bacterias por el método de pulsos de oxígeno («oxygen-pulse»: método que consiste en la introducción de cantidades pequeñas de oxígeno) y en los cloroplastos midiendo la traslocación de protones que sucede a un destello de luz. Recientemente, Martin D. Brand, Baltazar Reynafarje y Albert L. Lehninger, en la Johns Hopkins School of Medicine, han comunicado hallazgos experimentales en los que las mitocondrias transfieren más de seis protones por cada par de electrones que pasa del NADH al oxígeno. Si sus experimentos son correctos, será necesario modificar sustancialmente la teoría de Mitchell. Uno de nosotros (Hinkle) ha estudiado la traslocación de protones en vesículas formadas a partir de membrana mitocondrial y en vesículas de membrana artificial con enzimas respiratorios mitocondriales. La proporción $H^+/2e^-$ medida en estos sistemas concuerda con los hallazgos de Mitchell y Moyle, pero la controversia no está aún resuelta.

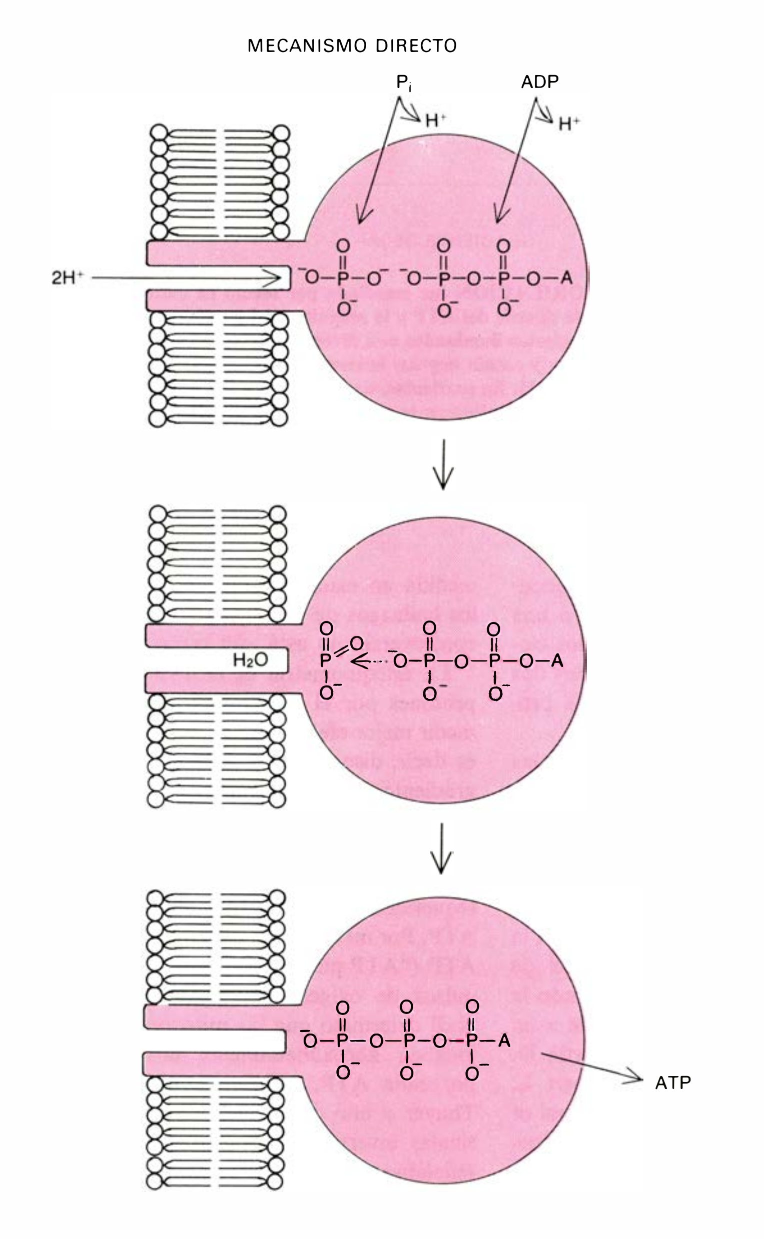
La estequiometría de la traslocación de protones por el complejo F_1-F_0 se puede medir mejor efectuando el proceso al revés, es decir, disociando el ATP para crear un gradiente de protones. Se mide entonces el número de protones transferidos por cada molécula de ATP disociada, lo que presumiblemente equivale al número de protones requeridos para sintetizar una molécula de ATP. Por medio de un método de pulsos de ATP ("ATP-pulse") análogo al método de pulsos de oxígeno ("oxygen-pulse'), Mitchell determinó que las mitocondrias traslocaban aproximadamente dos protones por cada ATP, hallazgo confirmado por Thayer y uno de nosotros (Hinkle) en vesículas invertidas de mitocondrias. Como señalábamos, los enzimas correspondientes de los cloroplastos, CF_1-F_0 parecen requerir tres protones por ATP, pero es difícil efectuar mediciones directas de esta proporción. Se han comunicado valores que varían de 2 a 4, y la proporción de 3 se dedujo indirectamente.

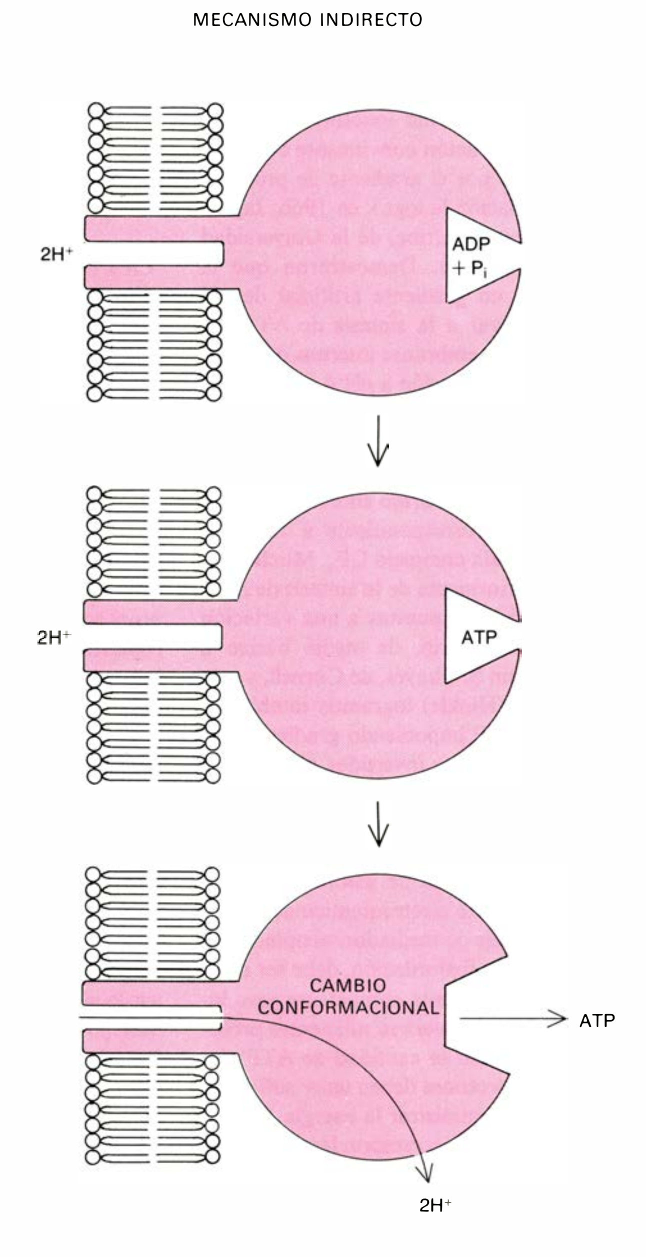
A partir de estas proporciones puede estimarse la estequiometría global de la fosforilación oxidativa y fotosintética. Si por cada par de electrones se transfieren en las

mitocondrias seis protones, y dos protones deben retornar a través de la membrana por cada ATP formado, entonces no pueden sintetizarse más que tres moléculas de ATP por cada par de electrones. Esta proporción no significa más que un límite superior, ya que se formarán menos moléculas de ATP si algunos de los protones son desviados a otros procesos, como el del transporte activo. La proporción global se ha determinado experimentalmente muchas veces midiendo el número de moléculas de ATP sintetizadas por cada átomo de oxígeno reducido a agua. En las mitocondrias los resultados experimentales tienden generalmente a indicar una proporción de tres moléculas de ATP por cada par de electrones que pasa del NADH al oxígeno. Algunos estudios recientes, no obstante, han sugerido que la proporción pudiera ser solamente de dos. Más adelante se discutirá un posible mecanismo basado en la necesidad de emplear energía para transportar el ADP y el ATP, que podría explicar esta baja proporción.

En los cloroplastos, en los que sólo son transferidos cuatro protones por cada par de electrones y se requieren tres protones por cada ATP sintetizado por el CF_1-F_0, la máxima estequiometría de la síntesis sería de cuatro moléculas de ATP por cada tres pares de electrones. La mayoría de las proporciones observadas son coherentes con este valor, aunque ocasionalmente han sido comunicadas proporciones mayores de 4/3.

A partir de estas estequiometrías puede trazarse un balance energético de la síntesis de ATP. Como hemos calculado anteriormente, el gradiente electroquímico de protones de los cloroplastos representa un almacenamiento de energía de unos 4,8 ki-





LOS POSIBLES MECANISMOS de la fosforilación que tiene lugar en el complejo F_1-F_0 pueden separarse en dos categorías, denominadas directa e indirecta. En el mecanismo directo *(a la izquierda)*, se une primero un ión fosfato y el ADP a la parte F_1 del complejo enzimático. Los protones se trasladan a través de un canal de la parte F_0 y atacan entonces uno de los oxígenos del fosfato, que es separado para formar una molécula de agua. Finalmente, un oxígeno del ADP ataca al átomo de fósforo dando lugar al ATP. La molécula de ATP se separa entonces del enzima. Es posible toda una variedad de métodos indirectos; un ejemplo de ellos se muestra aquí *(a la derecha)*. En el locus activo del enzima podría ser que el ADP y el fosfato inorgánico se combinaran espontáneamente sin la adición de energía libre. La molécula de ATP resultante quedaría, sin embargo, firmemente unida al enzima, y sería necesario aplicar energía para liberarla. La energía podrían suministrarla los protones, uniéndose al enzima en una posición distinta del locus activo. Los protones serían luego liberados a la solución por la parte F_1 de la membrana.

localorías por mol de protones. Dado que por cada ATP sintetizado tres protones retornan a través de la membrana, se dispone de 3 × 4,8, es decir, 14,4 kilocalorías para la fosforilación. Otras estimas sugieren que la máxima energía libre necesaria para la síntesis de ATP en los cloroplastos es de 14,5 kilocalorías por mol, con lo que el balance energético casi se equilibra.

Los requerimientos energéticos de la fosforilación mitocondrial son, en cierta forma, más complicados. Como se recordará, los protones que atraviesan la membrana mitocondrial tienen una energía de 5,3 kilocalorías por mol, pero sólo se consumen dos protones por cada ATP formado. La energía libre disponible es, consiguientemente, de 10,6 kilocalorías por mol, lo que resulta insuficiente, dado que se requieren unas 15 kilocalorías para la fosforilación. Hay, no obstante, una diferencia importante entre la fosforilación en los cloroplastos y en las mitocondrias, ya que en las últimas la reacción tiene lugar en el interior de la membrana interna. En este medio, la alta concentración de los reactivos (ADP y fosfato) y la baja concentración del producto (ATP) desplaza el punto de equilibrio de la reacción y reduce apreciablemente la energía libre requerida. En el interior de las mitocondrias, la síntesis de ATP sólo necesita unas 11 kilocalorías por mol, que queda dentro del intervalo de energía disponible que se ha calculado.

La reducción de los requerimientos de energía en la fosforilación mitocondrial no se salda sin un costo. El costo es la energía necesaria para concentrar el ADP y el fosfato en el interior de la cavidad y extraer el ATP. A partir de un estudio de las reacciones de transporte, Martin Klingenberg y sus colegas de la Universidad de Munich han llegado a la conclusión de que el contraflujo de iones ADP y ATP entraña el movimiento neto de cargas a través de la membrana. El intercambio de reactivos y productos podría, por tanto, tomar la energía del potencial de membrana. De forma similar, el fosfato es absorbido a cambio de iones OH^- y su transporte está consecuentemente dirigido por el gradiente de *p*H. El efecto combinado de estas transacciones es transportar un protón adicional por cada ATP sintetizado y exportado. Se gastan, pues, un total de tres protones por cada ATP que aparece en el exterior de la mitocondria, y la energía total disponible es de 3 × 5,3, es decir, 15,9 kilocalorías por mol. Con ello la estequiometría de la fosforilación mitocondrial ha variado: la proporción entre iones fosfato consumidos y átomos de oxígeno empleados se reduce de 3 a 2.

Si un gradiente de protones ha de ser un mediador satisfactorio entre el flujo de electrones y la fosforilación, no sólo debe proporcionar suficiente energía libre sino también debe ser cinéticamente competente. La velocidad con que se producen los protones debe ser, como mínimo, igual a la velocidad con que se consumen.

Thayer y uno de los autores (Hinkle) hicieron una rigurosa prueba de la teoría quimiosmótica comparando la velocidad de síntesis de ATP debida a la respiración en vesículas mitocondriales invertidas con la dirigida por un gradiente electroquímico de protones artificial. En la primera décima de segundo, el gradiente artificial condujo la síntesis de ATP más rápidamente de lo que lo hacía la oxidación del NADH, pero después la síntesis dirigida por el gradiente artificial disminuía a medida que iba decayendo el gradiente. Ello implicaba que la velocidad de fosforilación está controlada por la magnitud del gradiente. También se observó una diferencia de velocidad inicial en dos sistemas de vesículas con gradientes obtenidos naturalmente. Una de las suspensiones de vesículas tenía todos los reactivos necesarios para la fosforilación oxidativa a excepción del NADH, que se añadía en el último momento para dar comienzo a la reacción; en la otra suspensión, el componente que faltaba era el ADP. El sistema que se disparaba añadiendo NADH comenzaba mucho más despacio, presumiblemente porque se requería un cierto tiempo para que se formara el gradiente de protones. En el sistema en que faltaba ADP, la respiración podía establecer de antemano el gradiente y la fosforilación daba comienzo inmediatamente después de añadirse el ADP.

Más recientemente Boyer y sus colegas han mostrado que la velocidad de síntesis de ATP en cloroplastos expuestos a una transición de ácido a base es equivalente a la velocidad máxima obtenida con luz. A. R. Portis, Jr., y uno de nosotros (McCarty) encontramos que la velocidad de fosforilación en los cloroplastos depende muy directamente de la magnitud del gradiente de *p*H. De hecho, la velocidad es proporcional al cubo de la concentración interna de protones, un hallazgo que concuerda con las indicaciones de que pasan tres protones a través del complejo CF_1-F_0 por cada molécula de ATP que se forma.

Una de las técnicas bioquímicas más eficaces consiste en disgregar un complejo enzimático para ensamblar luego determinadas partes del mismo, de forma que pueda examinarse aisladamente el funcionamiento de cada parte. Este método puede aplicarse al estudio de la fosforilación oxidativa y fotosintética mediante la inclusión de determinadas proteínas en una membrana artificial.

En 1971, Yasuo Kagawa y Racker, que se habían trasladado a Cornell, desarrollaron un método para incorporar enzimas transferidores aislados a vesículas artificiales compuestas de fosfolípidos (las moléculas lipídicas que forman la matriz fundamental de todas las membranas biológicas). Las vesículas se obtenían mezclando los fosfolípidos y las proteínas seleccionadas con un detergente, y extrayendo el detergente por diálisis. Las moléculas se organizan espontáneamente en vesículas cerradas.

Las primeras vesículas fabricadas por Kagawa y Racker llevaban incorporado el complejo mitocondrial F_1-F_0. Estaban invertidas respecto de las mitocondrias normales; es decir, las prominencias F_1 se hallaban en la parte exterior en vez de situarse en la interior. En consecuencia, la disociación del ATP por el complejo F_1-F_0 transportaba los protones hacia dentro.

Desde entonces, uno de nosotros (Hinkle), con sus colaboradores, ha venido empleando el mismo método para estudiar cada uno de los tres segmentos de la cadena respiratoria mitocondrial por separado. Cada uno de los sistemas podía dirigirse añadiendo una combinación distinta de oxidantes y reductores. En cada caso, el flujo de electrones resultante estaba acoplado al transporte de protones con una estequiometría similar a la observada en las mitocondrias. Se lograba tanto un gradiente de concentración de protones como un potencial de membrana.

Racker y sus colaboradores han construido vesículas que contienen un complejo F_1-F_0 y, además, segmentos de la cadena respiratoria. Al preparar estas partículas hay que tener cuidado de que todos los componentes de una vesícula estén del mismo lado. Cuando todas las proteínas funcionales están orientadas correctamente, las vesículas son capaces de llevar a cabo la fosforilación oxidativa. Junto con Walther Stoeckenius, de la Universidad de California en San Francisco, Racker ha combinado también el complejo F_1-F_0 mitocondrial con una bomba de protones dependiente de la luz, denominada bacteriorodopsina, procedente de una halobacteria. Las vesículas híbridas muestran una fosforilación dependiente de la luz.

Estos sistemas reconstituidos ofrecen una prueba clara de que cada segmento de la cadena respiratoria transporta protones y de que un gradiente de protones por sí solo, cualquiera que sea la forma de generarlo, puede dirigir la fosforilación por el complejo F_1-F_0. Sugieren definitivamente que no sólo es necesario un gradiente de protones para la síntesis de ATP, sino también que es suficiente. La investigación de la teoría quimiosmótica puede, por tanto, pasar de la comprobación de los postulados fundamentales al examen detallado de los mecanismos.

Matemática combinatoria de la planificación

¿Cuál es el método óptimo de organizar un trabajo para terminarlo en el tiempo mínimo? Descubrimientos matemáticos y de ciencias de la computación revelan los méritos y deficiencias de los diversos métodos

Ronald L. Graham

Aparecen problemas de planificación en casi todas las áreas de la actividad humana. Para la misión Viking de exploración del planeta Marte fue preciso coordinar la actividad de más de 20.000 personas. Cumplir la cuota de producción diaria de una factoría automovilística puede exigir una muy estricta asignación de fuerza laboral y maquinaria. Incluso preparar una cena de menú muy variado puede plantear un problema de planificación nada trivial. Podría pensarse que existen algoritmos naturales, esto es, procesos por etapas bien definidas, para la construcción de planes altamente eficientes. No ocurre así, sin embargo. No se puede confiar en que métodos aparentemente lógicos de construcción de planificaciones den resultados igualmente buenos en situaciones diferentes. Por ejemplo, en ciertos casos, al aumentar el número de operarios dedicados a una tarea pueden producirse aumentos del tiempo total invertido en realizar un trabajo. Algunos de los más comunes e intuitivos procedimientos de planificación pueden dar lugar a resultados no esperados, e incluso, aparentemente paradójicos.

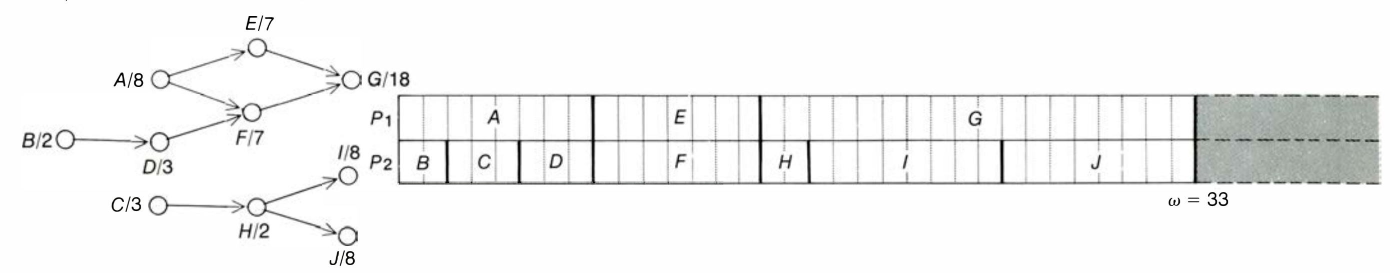
Durante los últimos años se han diseñado y estudiado con gran atención varios modelos matemáticos de procesos de planificación, en un esfuerzo encaminado a descubrir las causas de tales resultados anómalos. Fue necesario determinar cuán ineficientes podrían resultar distintos procedimientos de planificación, y cómo eliminar, o al menos, paliar al máximo, los rendimientos bajos. La investigación ha puesto de manifiesto que, en la mayoría de los casos, no debe esperarse que los algoritmos eficientes proporcionen resultados perfectos. No obstante, se han obtenido directrices de carácter general para rehuir los escollos latentes en problemas de planificación, y, también, para determinar soluciones aceptables de problemas concretos. Expondré aquí algunos resultados recientes de este carácter, aplicables a uno de los más básicos modelos de planificación. El modelo trata del comportamiento de conjuntos finitos de las magnitudes implicadas en los problemas de planificación. El análisis del modelo pone de relieve la productiva y fecunda interacción que suele darse entre las ciencias de computación y las matemáticas. En este caso se han apareado la teoría de algoritmos y la teoría combinatoria (en particular, el estudio de conjuntos finitos), generándose de esta forma importantes resultados en teoría de planificación.

Para comprender y profundizar en las dificultades de los problemas de planificación, el modelo ha de aislar las partes esenciales de las situaciones reales. El modelo básico está formado por un sistema de m procesadores idénticos, P_1..., P_m, y un conjunto de tareas A, B,..., T, que los procesadores han de realizar. (Los procesadores podrían ser operarios de una cadena de montaje o dispositivos físicos, como, por ejemplo, microprocesadores electrónicos.) A cada tarea T le está asignado un número positivo $\tau(T)$, que es el tiempo requerido por el procesador para ejecutar la tarea T, por lo que se llamará tiempo de ejecución de T. Aunque cada procesador es capaz de realizar cada una de las tareas, ningún procesador puede ejecutar más de una tarea por vez. Además, cuando un procesador empieza a realizar una tarea T, tiene que continuar ejecutándola hasta terminarla por completo.

Un algoritmo de planificación es el conjunto de reglas por las que se van asignando tareas a los procesadores individuales; cambiando el sistema de reglas se obtiene un algoritmo diferente. En el modelo, el orden en que se van eligiendo las tareas a realizar depende de dos factores: una colección de restricciones de precedencia y una lista de prioridad. Se dice que la tarea R es predecesora de la tarea S cuando, para comenzar con la tarea S, es imprescindible que R esté finalizada por completo. Las relaciones de esta clase se llaman restricciones de precedencia, y se denotan por $R \rightarrow S$. Una lista de prioridad L es una ordenación de las tareas de acuerdo con las preferencias del planificador: $L = (B,C,R,...)$. No es preciso que la lista sea coherente con las restricciones de precedencia que puedan existir. El modelo así obtenido suele llamarse modelo de planificación por lista de prioridad, pues para modificar las planificaciones que el modelo construye no hay más remedio que alterar la lista de prioridad.

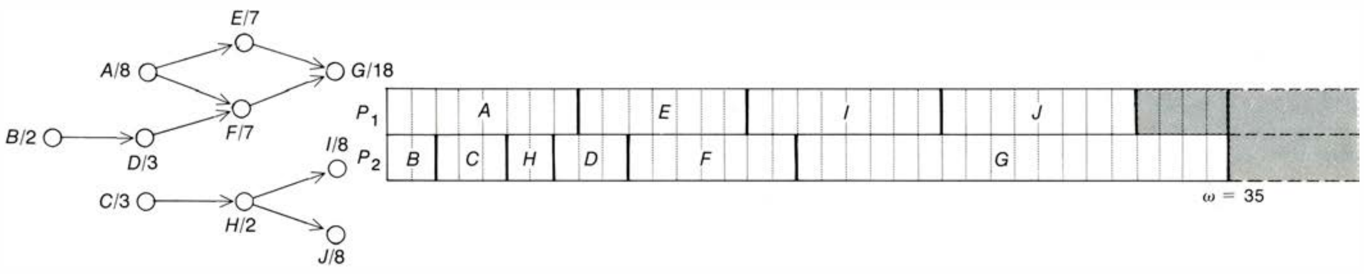
El modelo empieza a funcionar en el tiempo $t=0$ explorando todos los procesadores la lista de prioridad del conjunto de tareas a partir del principio, en búsqueda de "tareas disponibles". Se dice que una tarea está disponible o lista para ejecución si todas las tareas predecesoras han sido ejecutadas ya y ningún procesador ha comenzado a efectuarla. Evidentemente, en el tiempo $t=0$ las tareas disponibles son las que carecen de predecesoras. Si hay dos o más procesadores compitiendo por una misma tarea, ésta se le asigna al procesador P_i que tenga menor número índice i. Cuando un procesador no puede encontrar ninguna tarea disponible, deja de explorar la lista de prioridad y queda inactivo. El procesador permanece inactivo hasta que en algún lugar del sistema se termine de ejecutar una tarea. Esta tarea po-

L = (*A*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*)

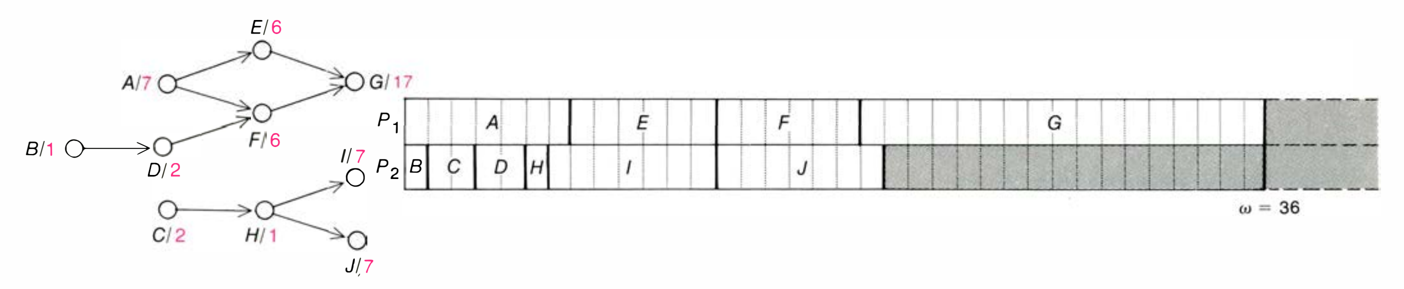


EN UN MODELO DE PLANIFICACION se aíslan los componentes esenciales de las situaciones de planificación reales. El modelo puede utilizarse para estudiar el comportamiento de algoritmos, descripciones etapa por etapa de un procedimiento, en orden a constituir planificaciones eficientes. El modelo fundamental está constituido por un conjunto de procesadores idénticos y un conjunto de tareas que los procesadores han de realizar ajustándose a ciertas reglas. A cada tarea *T* le está asignado un tiempo de ejecución $\tau(T)$, igual al tiempo necesario para ejecutarla. En el ejemplo mostrado en esta ilustración hay 10 tareas, *A*, *B*,..., *J* y dos procesadores, P_1 y P_2. Las tareas a realizar están relacionadas por diversas restricciones de precedencia. Se dice que la tarea *Q* es predecesora de la tarea *S* si la tarea *Q* ha de estar terminada para que sea posible comenzar *S*. Esta relación se llama restricción de precedencia, y se denota $Q \rightarrow S$. Un algoritmo de planificación es el conjunto de reglas por las que se van asignando tareas a los procesadores. En el modelo, para la distribución de tareas se utiliza una lista de prioridad, que es una lista de las tareas ordenada según el criterio del planificador. No es necesario que la lista de prioridad sea coherente con las restricciones de precedencia que pueden existir. La lista de prioridad *L* de este ejemplo se muestra arriba, a la izquierda; el grafo *(abajo, a la izquierda)* permite visualizar las tareas *(círculos en blanco)* y las restricciones de precedencia entre ellas *(flechas)*. Cada tarea del grafo, *T*, está rotulada con su nombre y tiempo de ejecución: $T/\tau(T)$. La distribución de tareas comienza con todos los procesadores explorando desde el principio la lista de prioridad en busca de tareas disponibles esto es, de tareas cuyas restricciones de precedencia hayan sido satisfechas. Una vez localizada una tarea disponible, se le asigna al procesador P_i desocupado de menor subíndice. En este caso, la primera tarea disponible de la lista *L* es la tarea *A*, y se le asigna al procesador P_1. Tras asignarle tarea a un procesador, los todavía desocupados prosiguen la exploración de la lista. El diagrama de tiempos de la derecha muestra el plan-horario, o asignación de tiempos de procesamiento, determinado por las reglas del modelo. Las zonas sombreadas representan tiempos muertos, donde un procesador permaneció inactivo. En este modelo, un procesador solamente puede permanecer inactivo si no existen tareas disponibles. Por ello, cuando acaba de ejecutarse una tarea, todos los procesadores inactivos comienzan a explorar de nuevo la lista. El plan concluye cuando todas las tareas han sido ejecutadas.

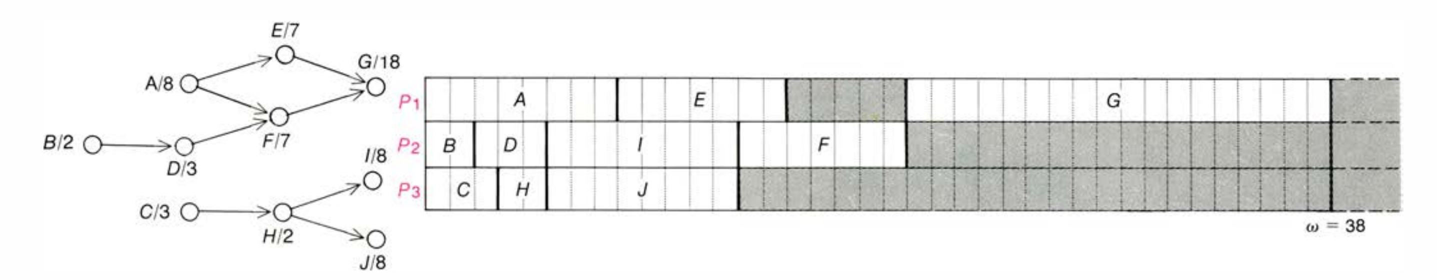
L = (*A*, *B*, *C*, *H*, *D*, *E*, *F*, *G*, *I*, *J*)



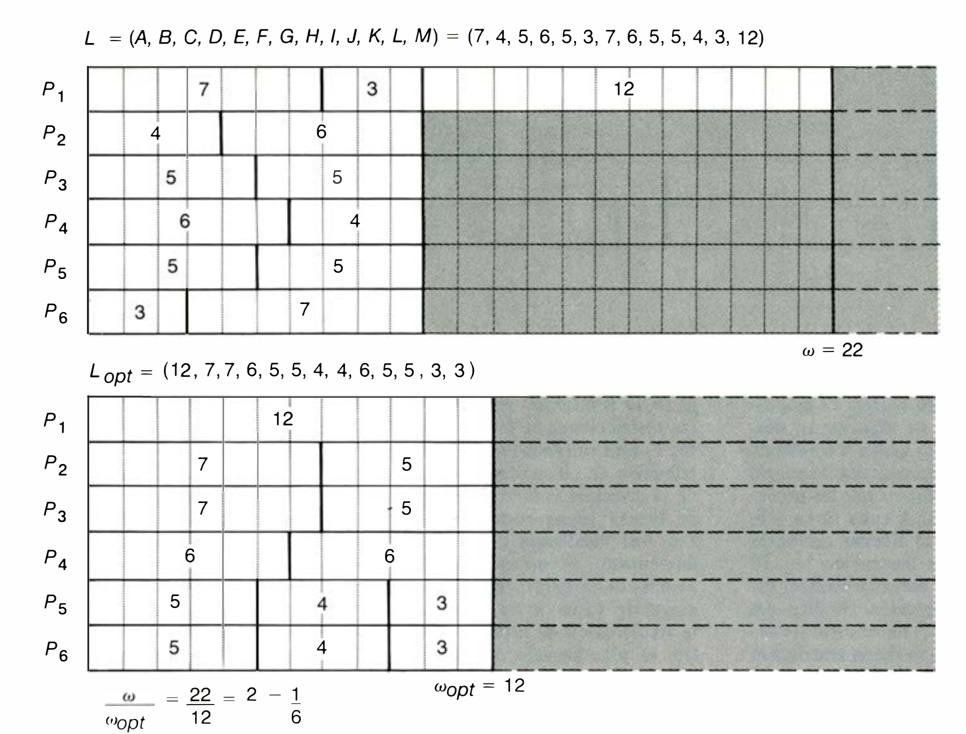
L = (*A*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*)



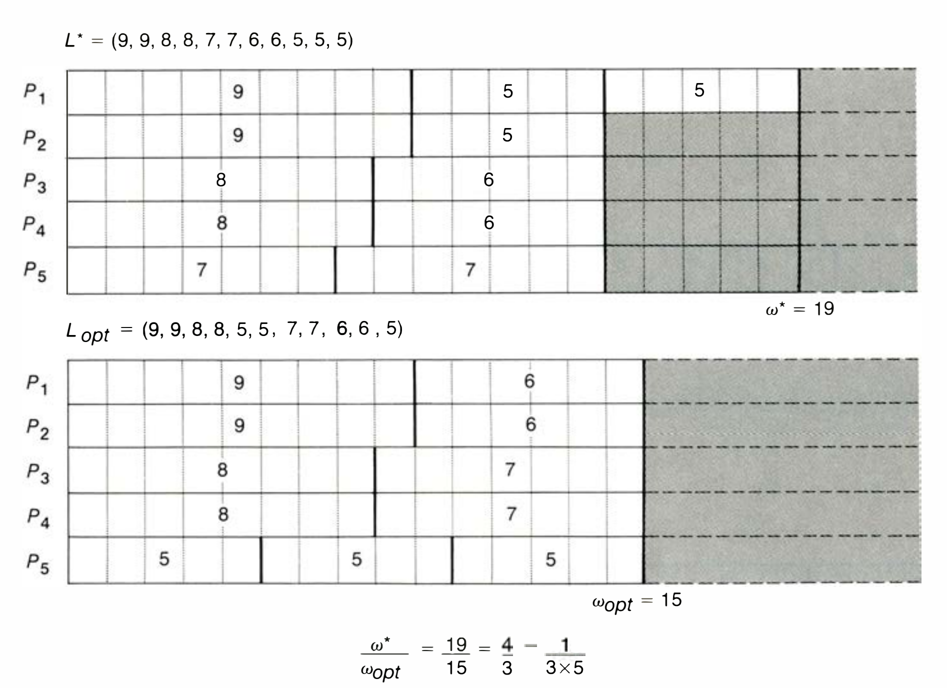
L = (*A*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*)



MODIFICAR LOS PARAMETROS del modelo fundamental de planificación puede tener efectos imprevisibles, reordenando la lista de prioridad *L* del problema de la ilustración superior, el tiempo de terminación aumenta a 35. Si todos los tiempos de ejecución se reducen en 1 *(centro)*, se obtiene un plan de trabajo aún más sorprendente: el tiempo de terminación aumenta a 36. Además, con estos tiempos de terminación menores el tiempo de terminación es siempre 36, sea cual fuere la lista de prioridad. Ni añadiendo un tercer procesador al sistema original *(abajo)* se rebaja el tiempo de terminación. Por el contrario aumenta hasta 38,y pasa igual con todas las listas de prioridad.



UN PLAN ES OPTIMO para un conjunto dado de tareas si su tiempo de terminación es el mínimo que puede obtenerse reordenando convenientemente la lista de prioridad de las tareas. En el problema de planificación que aquí se muestra no hay restricciones de precedencia entre las tareas; por ello la asignación de tareas a los procesadores se produce en el mismo orden con que figuran en la lista de prioridad L. El tiempo de terminación del plan de trabajo resultante *(arriba)* es $\omega = 22$; durante su realización muchos procesadores están inactivos por largos períodos de tiempo. Una diferente ordenación de las tareas produce un tiempo de terminación más corto *(abajo)*. Este plano es óptimo, porque ningún procesador se encuentra nunca inactivo antes del tiempo de terminación $\omega_{opt} = 12$. La razón de los dos tiempos de finalización es la máxima posible, pues se ha podido demostrar que ω/ω_{opt} nunca es mayor que $2-1/m$, siendo m el número de procesadores utilizados. En la figura se pueden reconocer las tareas por la duración de su ejecución. Al no existir restricciones de precedencia (tareas independientes) no hay peligro de confusión.



EN EL ALGORITMO DE LISTA DECRECIENTE que aquí se presenta, las tareas más largas se ejecutan lo antes posible, con lo que hay menos riesgo de que hacia el final del plan de trabajo algunos procesadores hayan de permanecer inactivos mientras otros estén operando todavía. Con este algoritmo se forma una lista de prioridad L^* situando las tareas en orden de tiempos de ejecución decrecientes. Cuando hay restricciones de precedencia entre las tareas, el algoritmo puede tener rendimientos muy bajos, pero cuando las tareas son independientes, el algoritmo de lista decreciente garantiza resultados bastante por debajo de la costa $2-1/m$. En realidad, cualquiera que sea la lista de tareas, L, ω^*/ω_{opt} (razón de los tiempos de terminación de un conjunto de tareas con el plan de lista decreciente al tiempo óptimo de terminación para ese conjunto) nunca es mayor que $4/3-1/3\ m$. En el presente ejemplo de cinco procesadores se alcanza esta cota superior.

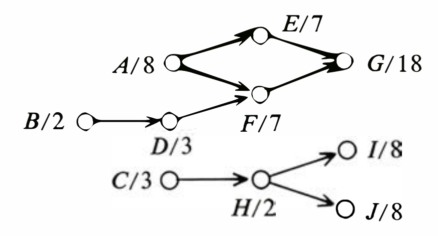
dría ser predecesora de otras tareas, que pasarán a estar disponibles a su terminación. Así pues, siempre que se concluye una tarea, todos los procesadores inactivos comienzan instantáneamente a explorar la lista de prioridad desde el principio. La planificación terminará cuando se haya finalizado la última tarea de la lista; el tiempo de terminación se denota ω. (Para simplificar, el tiempo consumido en explorar la lista no se incluye en el tiempo de funcionamiento del modelo.)

El modelo, que describe un caso de planificación típico, fue estudiado al objeto de hallar respuesta para las siguientes cuestiones: ¿De qué forma depende el tiempo de terminación de la lista de prioridad elegida? ¿Cómo depende de las restricciones de precedencia? ¿Qué relación tiene con los tiempos de ejecución de las tareas? ¿Cómo depende del número de procesadores? En particular, ¿cómo se podrá elegir la lista de prioridad para hacer mínimo el tiempo de terminación? Dicho de otra forma, el modelo sirve para evaluar el algoritmo de planificación, teniendo en cuenta su rendimiento en distintas condiciones, por ejemplo, con procesadores más rápidos, menor número de tareas, etcétera.

El modelo no es aplicable a todos los casos de planificación. Por ejemplo, no se toman en cuenta los aspectos probabilísticos de la confección de un plan de trabajo en el que, por ejemplo, los tiempos de ejecución de las tareas no sean fijos, sino aleatorios con cierta distribución de probabilidad. Además la única medida de rendimiento de un algoritmo que el modelo toma en cuenta es el tiempo de terminación, cuando en realidad puede haber varias otras más. Por otra parte, el modelo cuenta con cierto número de características compartidas por muchas situaciones reales. Los procesadores del modelo podrían ser mecanógrafas de una compañía, y las tareas podrían consistir en un conjunto de informes que es preciso pasar a máquina. Aunque el informe del presidente de la compañía podría muy bien encabezar la lista de prioridad, su informe dependerá probablemente de los resultados de informes preparados por subordinados suyos; es decir, tales informes serían entonces predecesores del informe presidencial. Los procesadores del modelo podrían representar también miniordenadores de un sistema de computación con multiprocesos efectuando las diversas subrutinas de un programa complejo. Aunque el modelo es bastante sencillo, tiene suficiente estructura para poner de relieve toda la gama de dificultades

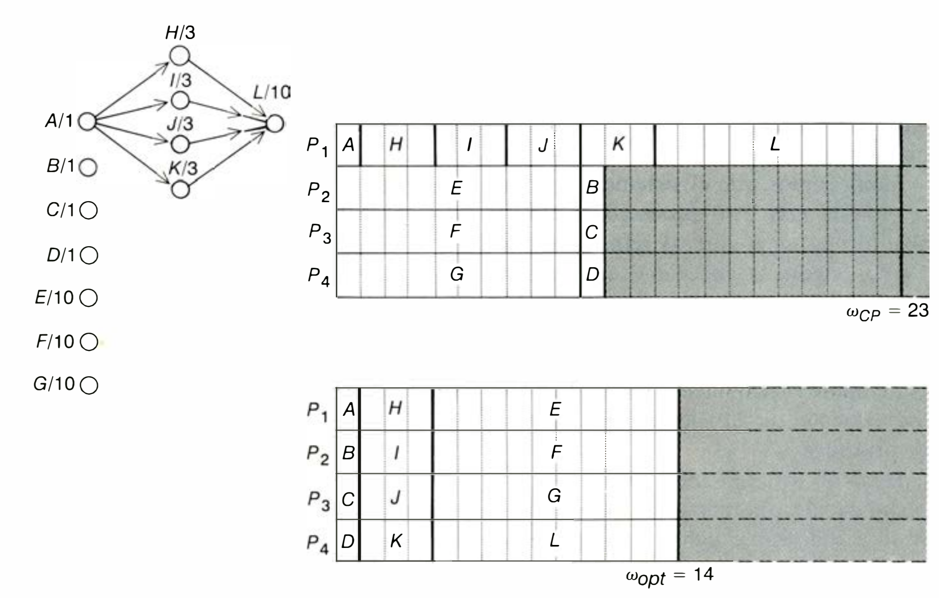
que se pueden encontrar en problemas combinatorios generales de preparación de planes de trabajo. Un ejemplo ayudará a mostrar que así ocurre.

Supongamos que haya dos procesadores, P_1 y P_2, y 10 tareas a ejecutar, ordenadas en una lista: $L=(A, B, \ldots, J)$. Para visualizar las restricciones de dependencia entre tareas y sus respectivos tiempos de ejecución resulta conveniente preparar un diagrama en el que cada círculo en blanco representa una tarea T, cada flecha representa una restricción de precedencia y las tareas llevan la leyenda $T/\tau(T)$:

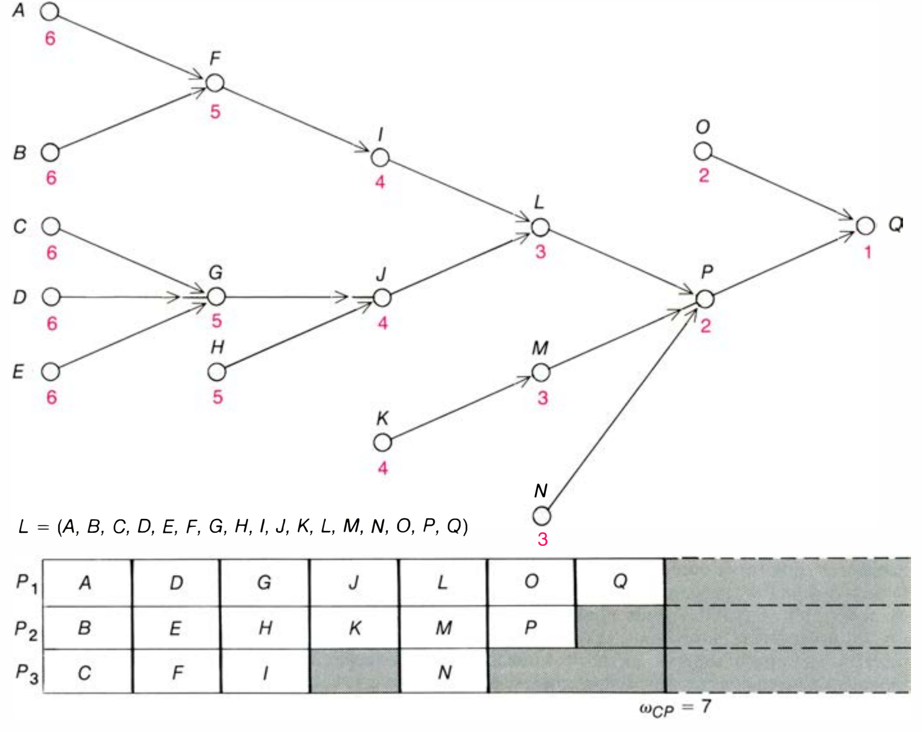


En el instante $t=0$ los dos procesadores comienzan a explorar L, e inmediatamente llegan a la tarea A, la cual, por carecer de predecesoras, está lista para ejecución. De acuerdo con las reglas del modelo, la tarea A se le asigna al procesador P_1, pues tiene menor subíndice. La siguiente tarea disponible es B, que le es asignada a P_2. En el tiempo $t=2$ se termina la ejecución de la tarea B, y P_2 puede volver a explorar la lista L; al hallar disponible la tarea C, comienza a ejecutarla. El proceso continúa de esta forma hasta el tiempo $t=33$ en que se completa el conjunto de 10 tareas *(véase la ilustración superior de la página 77.)* Como ningún procesador estuvo inactivo en ninguna ocasión antes de $t=33$, la lista L es óptima en el sentido de que, independientemente de la forma en que se reordenen las tareas, no será posible acabarlas todas ellas en una cantidad menor de tiempo.

No se conseguirá rebajar el tiempo $\omega=33$ cambiando la ordenación de la lista. Pero, ¿qué ocurrirá si se modifican otros parámetros del modelo, como, por ejemplo, los tiempos de ejecución de las distintas tareas o el número de procesadores? Se obtienen resultados sorprendentes *(véase la ilustración inferior de la página 77)*. Si se disminuyen en una unidad todos los tiempos de ejecución, el tiempo de terminación sube a 36. Para explicar este extraño resultado, podríamos caer en la tentación de postular sin más que la lista de prioridad elegida, L, es sumamente inadecuada para los nuevos tiempos de ejecución, y que con una lista de prioridad mejor elegida el tiempo de terminación se reduciría.



LA PLANIFICACION POR CAMINOS CRITICOS es uno de los algoritmos más utilizados, aunque su rendimiento no siempre es bueno. En un plan de trabajo por caminos críticos, como el que se muestra arriba, la asignación de tareas a los procesadores se hace según la urgencia relativa de las mismas, esto es, de acuerdo con la longitud (que es la suma de los tiempos de ejecución) de las diferentes cadenas de procesamiento que cada tarea encabeza en la parte todavía no ejecutada del diagrama de restricciones de precedencia. Las cadenas más largas del diagrama son los llamados caminos críticos, porque dichas cadenas son las que con mayor probabilidad provocarán embotellamientos en la ejecución del conjunto de tareas. En la planificación por caminos críticos se elige siempre como próxima tarea a ejecutar la que esté encabezando el camino crítico de ese momento. En este ejemplo, en el tiempo $t=0$ hay cuatro caminos críticos, todos de longitud 14: $A \to H \to L$, $A \to I \to L$, $A \to J$ y $A \to K \to L$. Así pues, la primera tarea a ser ejecutada es la tarea A, que se le asigna al procesador P_1 porque P_1 tiene menor subíndice que P_2, P_3 o P_4. En algunos casos, las planificaciones por caminos críticos son extraordinariamente eficientes, pero en este caso particular, el tiempo de terminación ω_{CP} del plan de caminos críticos es casi el doble de ω_{opt}, tiempo.



LAS PLANIFICACIONES POR CAMINOS CRITICOS SON OPTIMAS en aquellos problemas donde los tiempos de ejecución de las tareas sean idénticos y el sistema de restricciones de precedencia sea arborescente, es decir, que ninguna tarea tenga más de una sucesora inmediata. Como en este caso particular todos los tiempos de ejecución son iguales (por ejemplo, 1), la planificación por caminos críticos puede realizarse asignándole a cada tarea T un número $L(T)$ igual al número de tareas de la cadena más larga encabezada por T. El número $L(T)$ se llama "nivel de T", y el plan se elabora eligiendo siempre como próxima tarea a realizar la de máximo nivel entre las disponibles. En el ejemplo de tres procesadores aquí expuesto se han indicado los niveles en color. En el tiempo $t=0$ las tareas disponibles son A, B, C, D, E, H, K, N y O. Las tareas A, B, C, D y E tienen nivel máximo, 6, por lo que se elige A para ser ejecutada en primer lugar. T. C. Hu ha demostrado que el algoritmo de niveles crea siempre algoritmos óptimos en este caso particular; por consiguiente, $\omega_{CP}=7$ es el tiempo óptimo de finalización para este problema concreto.

Sin embargo, experimentando un poco se observa que con los nuevos tiempos de ejecución el tiempo de terminación nunca es inferior a 36, se ordenen como se ordenen las tareas.

Supongamos que el sistema primitivo se amplía con otro procesador. No es absurdo suponer que al reforzar en un 50 por ciento la capacidad de procesamiento del sistema se obtendrá una reducción comparable del tiempo de terminación. En este ejemplo, sin embargo, el tiempo de terminación es 38, independientemente de la elección de la lista de prioridad.

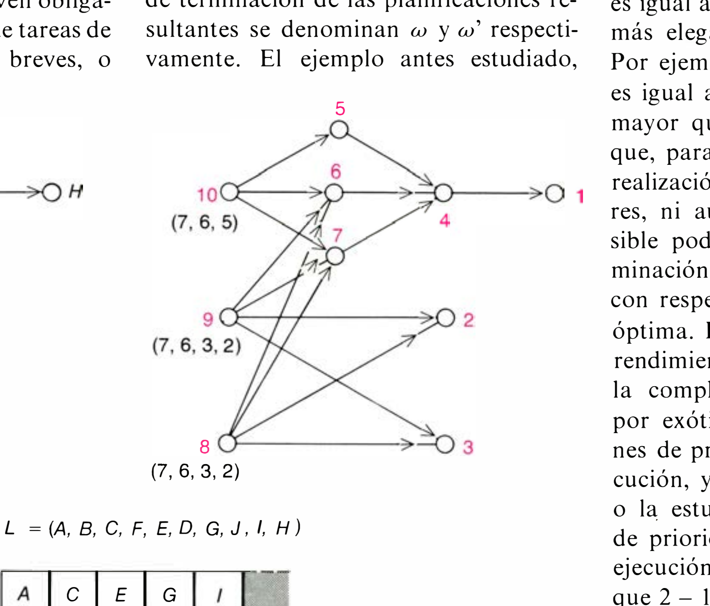
Dos son los elementos del modelo de planificación que provocan tan imprevisible e indeseable comportamiento. En primer lugar, no se permite que ningún procesador permanezca inactivo si el sistema contiene alguna tarea disponible para él; y, en segundo, una vez que el procesador comienza a ejecutar una tarea, no se le permite interrumpirla hasta que la termine por completo. En consecuencia, los procesadores se ven obligados a comenzar la ejecución de tareas de relativa poca monta (tareas breves, o ligadas a pocas restricciones de precedencia), y en cuanto las han comenzado no pueden interrumpir su ejecución para afrontar otras tareas más urgentes que van quedando disponibles. Evidentemente, podría modificarse la estructura del modelo, pero la verdad es que en su forma presente refleja muchas situaciones reales. Así pues, es interesante conocer cuánto efecto puedan tener estas peculiaridades del modelo en las planificaciones que a partir del mismo se construyan. Esta cuestión tiene una respuesta inesperadamente simple.

Imaginemos que un conjunto de tareas se ejecuta dos veces: en una ocasión con m procesadores, una lista de prioridad determinada, un conjunto de restricciones de precedencia y un conjunto de tiempos de ejecución, y en otra, con m' procesadores, una lista diferente, un conjunto más débil de restricciones de precedencia y un conjunto de tiempos de terminación más reducidos. Los tiempos de terminación de las planificaciones resultantes se denominan ω y ω' respectivamente. El ejemplo antes estudiado, con dos procesadores y un conjunto de 10 tareas, mostró que la razón ω'/ω de los tiempos de terminación puede ser mayor que 1, esto es, que mejorando los valores de los parámetros del modelo puede que inevitablemente se produzca un incremento del tiempo de terminación. Existe, sin embargo, un límite máximo para dicha razón y, por tanto, un tope máximo para los efectos adversos de la estructura del modelo. Hace algunos años pude demostrar que ω'/ω es siempre menor o igual que $1+(m-1)/m'$, esto es, que $\omega'/\omega \leq 1+(m-1)/m'$. Además, existen ejemplos en los que ω'/ω es igual a la cota superior $1+(m-1)/m'$, por lo que dicha cota puede mejorarse poniendo en su segundo miembro una expresión menor. (Si se utilizase una cota superior más pequeña, se obtendrían contradicciones en ciertos ejemplos.)

En aquellos casos en que se comparan las planificaciones producidas por diferentes listas de prioridad el número de procesadores es el mismo. Así pues, m es igual a m', y la desigualdad adopta su más elegante expresión: $\omega'/\omega \leq 2-1/m$. Por ejemplo, si hay tres procesadores m es igual a 3, por lo que ω'/ω nunca será mayor que 5/3; resultado que significa que, para todo conjunto de tareas cuya realización se encargue a tres procesadores, ni aún utilizando la lista peor posible podrá aumentar el tiempo de terminación en más de un $66\frac{2}{3}$ por ciento con respecto al determinado por la lista óptima. La cota $2-1/m$ es garantía de rendimiento: garantiza que sea cual sea la complicación del sistema de tareas, por exóticas que resulten las restricciones de precedencia o los tiempos de ejecución, y sin que importe la inteligencia o la estupidez con que se elija la lista de prioridad, la razón de los tiempos de ejecución ω'/ω nunca podrá ser mayor que $2-1/m$.

Es evidente que, a pesar de la acotación anterior, la elección de una buena o una mala lista de prioridad puede tener gran trascendencia en lo que a tiempo de terminación se refiere. El sentido común sugiere que las mejores listas podrían encontrarse entre las que sitúan al principio las tareas de tiempos de ejecución más largos. En tal caso, hacia el final del plan de trabajo solamente se irían sumando tiempos relativamente pequeños de utilización de los procesadores, y habría menor probabilidad de que algunos de ellos tuvieran que permanecer inactivos en espera de que otros acaben de realizar sus tareas.

Uno de los algoritmos de planificación más generalizados consiste en preparar la lista de prioridad L^* situando las tareas por orden de tiempos de eje-



EL ALGORITMO *CG* produce planificaciones óptimas cuando todas las tareas tienen iguales tiempos de ejecución y solamente dos procesadores deben ejecutarlas. En este algoritmo se le asignan a cada tarea números de prioridad *(en color)* de forma que aquellas tareas que encabecen largas cadenas de procesamiento o que tengan muchas sucesoras posean mayor prioridad. Antes de poder aplicar prácticamente el algoritmo *CG* es necesario depurar el diagrama de las restricciones de procedencia superfluas. (Por ejemplo, si $A \to B$ y $B \to C$ entonces debe eliminarse $A \to C$.) Se muestra, en la parte superior izquierda, un diagrama ya reducido. En el diagrama se han omitido los tiempos de ejecución, por ser todos iguales, por ejemplo, a 1. El algoritmo *CG* comienza numerando las tareas, como se muestra arriba, a la derecha. Primero se le asigna el número 1 a alguna tarea que carezca de sucesoras; si hubiese más de una se las numeraría 2, 3, y así sucesivamente. En la ilustración, la tarea *H* que no tiene sucesoras, ha sido numerada con un 1, y las tareas *I, J* han recibido, respectivamente, los números 2 y 3. A continuación, para cada tarea con la propiedad de que todas sus tareas hayan sido numeradas ya se forma una sucesión con los números asignados a todas sus sucesoras. La tarea cuya sucesión de números ocupe el primer lugar por orden "lexicográfico" entre las sucesiones construidas hasta ese momento será la que reciba el siguiente número. En el orden lexicográfico 5, 3, 2 va delante de 6, 1; de 5, 4, 3; de 5, 3, 2, 1, etcétera. Por ejemplo, la sucesión 7, 6, 3, 2, que es la sucesión asociada a las tareas *B* y *C* de la ilustración, va antes que la sucesión 7, 6, 5, asociada a la tarea *A*. Por consiguiente, a las tareas *B* y *C* se les asignan, respectivamente, los números 8 y 9, el 10 se asigna a la tarea *A*. Cuando todas las tareas han sido numeradas se forma la lista de prioridad *L* situando en ellas las tareas por orden decreciente de números asignados. En el diagrama de tiempos que se encuentra reproducido en la parte inferior de esta ilustración se nos muestra que el plan *CG* de trabajo es óptima para las tareas *A, B,... H.*

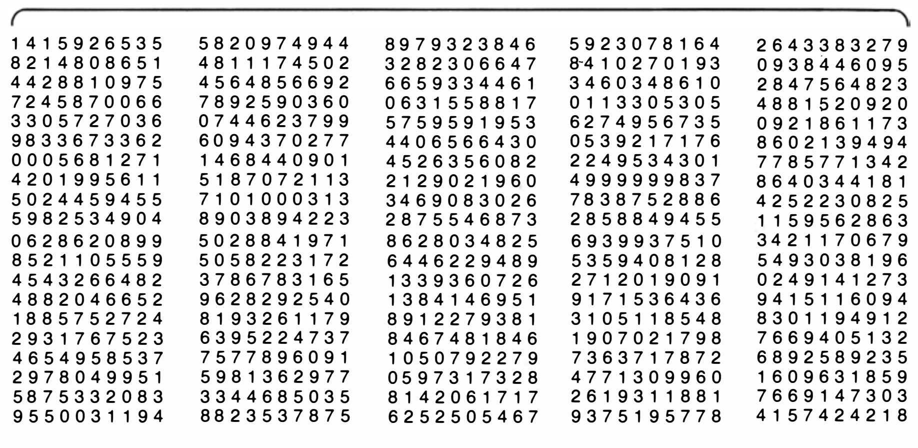
cución decrecientes, y ejecutarla después según las reglas del modelo. El tiempo de ejecución de este algoritmo por lista de tiempos decrecientes se denota ω^* *(véase la ilustración inferior de la página 78)*. ¿Cuál es el grado de bondad de este algoritmo? Dicho de otra forma, ¿cuánto se aproxima ω^* a ω_{opt}, que sería el tiempo óptimo (mínimo) de terminación del trabajo? En el caso de que existan restricciones de precedencia entre las tareas, puede ocurrir que este algoritmo confeccione el peor plan de trabajo posible, o lo que es igual, que la razón de ω^* a ω_{opt} alcance la cota $2 - 1/m$.

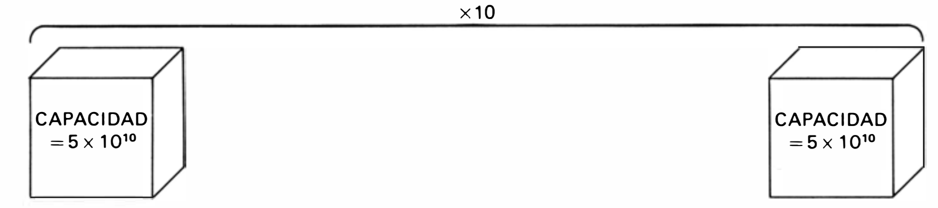
Cuando no hay restricciones de precedencia de distribución de tareas a los procesadores se hace en el mismo orden dado por la lista de prioridad. Se dice entonces que las tareas son independientes, habiéndose demostrado que en tal caso el algoritmo de lista decreciente da siempre resultados muy por bajo de la cota $2-1/m$. De hecho, para un conjunto de tareas independientes, la razón ω^*/ω_{opt} nunca supera el valor $4/3 - 1/3m$, que es significativamente menor que $2 - 1/m$ cuando m se hace grande. Como existen casos en que la razón de tiempos de ejecución alcanza la cota expresada, no es posible mejorar la acotación anterior. La desigualdad $\omega^*/\omega_{opt} \leq 4/3 - 1/3m$ garantiza que al aplicar el algoritmo de lista decreciente a un conjunto de tareas independientes nunca se confeccionará un plan de trabajo cuyo tiempo de terminación supere al óptimo en más de un 33⅓ por ciento.

Hasta el momento no se han hallado algoritmos de planificación de conjuntos de tareas con restricciones de precedencia, cuya eficiencia pueda compararse a la que presenta el algoritmo de lista decreciente en el caso de tareas independientes. En realidad no se conoce ningún procedimiento eficiente para construir listas de tareas interdependientes cuyos correspondientes planes de trabajo garanticen, con tres o más procesadores, tiempos de terminación menores o iguales que $2 - \epsilon$ veces el tiempo óptimo, para algún valor positivo de ϵ. La desigualdad $\omega'/\omega \leq 2 - 1/m$ muestra, no obstante, que el plan correspondiente a la lista peor posible tiene un tiempo de terminación inferior al doble del óptimo. Así pues, es evidente que habrá que progresar mucho en esta dirección.

La discusión anterior plantea una cuestión obvia: ¿Por qué dedicar tanto esfuerzo a crear planes de trabajo buenos, pero no óptimos? ¿Por qué no intentar hallar el plan óptimo para cada colección de tareas dada? Uno de los posibles métodos de hacerlo sería examinar todas las posibles planificaciones y elegir la de menor tiempo de terminación. Esta solución, que podríamos llamar "de fuerza bruta", no tiene más que un inconveniente: cuando el número de tareas es algo crecido el número de posibles listas de prioridad (y de correspondientes programaciones del trabajo) es enorme, y crece tan explosivamente que no hay esperanza de poder examinar ni siquiera una pequeña fracción del total. Si se tratase de un conjunto de n tareas, el conjunto de posibles listas constaría de $n!$ listas diferentes, es decir, $n(n-1)(n-2)\ldots1$, que incluso para valores relativamente pequeños de n resulta ser un número enorme. Por ejemplo, si se tratase de 20 tareas, e incluso si un ordenador pudiese examinar un millón de ellas por segundo, necesitaría más de 70.000 años para comprobar la totalidad de las 20! listas.

El número $n!$ de listas posibles (y, por tanto, el número de operaciones y el tiempo requerido por el ordenador para examinar los planes de trabajo) es función exponencial del número n de tareas. Las funciones exponenciales aumentan muy rápidamente al crecer el valor de la variable n. Una función polinómica, por ejemplo, n^2, no crece tan rápida y explosivamente al crecer n. Solamente sería práctico insistir en la búsqueda de planes óptimos de trabajo si se pudiera hallar un algoritmo para determinarlos en el que el número de pasos de cómputo a realizar creciera como función polinómica del número de tareas.





EN EL PROBLEMA DE EMBALAJE hay un conjunto de artículos, o pesos, y una colección de cajas idénticas capaces de contener un peso máximo fijo. El problema consiste en embalar todos los pesos del conjunto en el número mínimo de cajas. La dificultad de obtener soluciones exactas de los problemas de embalaje queda de manifiesto en este ejemplo. El total de los cien pesos indicados arriba suma 5×10^{11}. ¿Pueden alojarse los pesos en 10 envases de capacidad 5×10^{10} cada uno? El número de posibles modos de empaquetar este grupo relativamente reducido de pesos es tan enorme que incluso si se dispusiera de toda la capacidad de computación del mundo, resultaría improbable que pudiera hallarse respuesta a la pregunta anterior. La mayoría de los problemas de planificación son de complejidad similar; muchos se diseñan con la intención de obtener empaquetamientos con los que sólo puede garantizarse una aproximación razonable al caso óptimo.

Parece altamente improbable que llegue a descubrirse un algoritmo tal. Tan triste panorama es resultado del fundamental trabajo de Stephen A. Cook, de la Universidad de Toronto, quien introdujo en 1971 la noción de problema *NP*- completo, esto es, de problema de tiempo no determinísticamente polinómico y completo (véase "La eficiencia de los algoritmos", por Harry R. Lewis y Christos H. Papadimitriou, en INVESTIGACION Y CIENCIA, marzo, 1978). Se sabe ahora que centenares de problemas famosos, por no ser computacionalmente manejables, pertenecen a la clase de problemas *NP*-completos. Esta clase de problemas *NP*-completos tienen dos importantes propiedades. Primera, todos los métodos, o algoritmos, para hallar soluciones generales de tales problemas que se conocen hasta el presente exigen tiempos de ejecución de crecimiento exponencial, y, por consiguiente, son sumamente ineficientes. En segundo lugar, si uno solo de los pro-

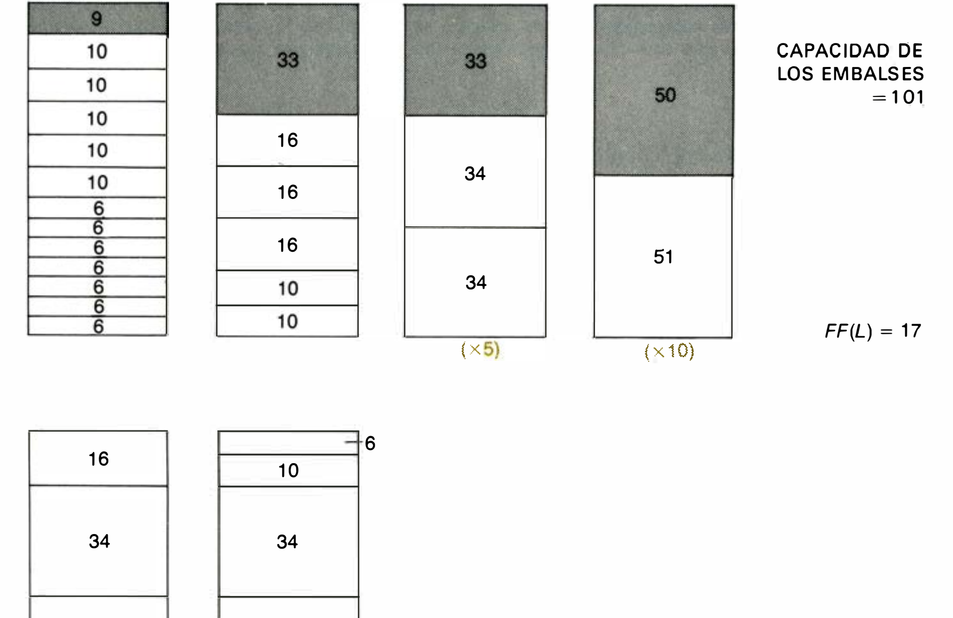
blemas NP-completos admitiera una solución eficiente, esto es, cuyo tiempo de ejecución crezca como las funciones polinómicas, entonces todos admitirían también soluciones eficientes. Aunque todavía no haya sido demostrado, parece sumamente probable que la dificultad de hallar procedimientos eficientes de resolver estos problemas sea intrínseca a los problemas NP-completos; verosímilmente, tales procedimientos no pueden existir.

La mayor parte de los problemas de planificación son NP-completos. De hecho, incluso en el caso relativamente simple de que no existan restricciones de precedencia, y aún con tan sólo dos procesadores, el problema resultante es NP-completo. Los descubrimientos relativos a la NP-completitud han alterado totalmente el rumbo de los trabajos de investigación sobre planificación. Los primeros esfuerzos se encaminaron a descubrir soluciones óptimas, exactas, de tales problemas, pero en la actualidad casi toda la atención se ha concentrado en la dirección, más fructífera, de hallar fácilmente soluciones aproximadas, o sea, determinar métodos eficientes que garanticen resultados cercanos a los óptimos. Como ejemplo de esta nueva línea de trabajo se tiene el algoritmo de lista decreciente, que garantiza tiempos de terminación no superiores al $33^1/_3$ por ciento de incremento sobre el óptimo.

El algoritmo de lista de tiempos decrecientes consigue un tiempo de terminación próximo al óptimo realizando cierta cantidad de trabajo. Las n tareas a ejecutar han de ser dispuestas en una lista de tiempos decrecientes, antes de distribuirse entre los procesadores. Tal clasificación puede llevarse a cabo en una cantidad de tiempo, o equivalentemente, en un número de etapas de cómputo, proporcional a $n \log_2 n$. Al crecer n la función $n \log_2 n$ crece sólo un poco más rápidamente, y, por tanto, el algoritmo es aceptablemente eficiente.

Trabajando un poco más se pueden obtener planificaciones todavía más cercanas a la óptima que la construida con el algoritmo de lista decreciente. Supongamos un sistema con dos procesadores. Un método de obtención de algoritmos perfeccionados consiste en seleccionar, para cierto entero k, las $2k$ tareas más largas, construir la planificación óptima para estas tareas y programar arbitrariamente las restantes. Si el tiempo de terminación de la planificación así obtenida se denota ω_k, puede demostrarse que $\omega_k/\omega_{opt} \leq 1 + 1/(2k+2)$. Con un conjunto de n tareas el procedimiento completo puede prepararse en no más de $2kn + 2^{2k}$ operaciones. (El término $2kn$ corresponde a la separación de las $2k$ tareas más largas, y el término 2^{2k} del examen exhaustivo de todas las posibilidades de ejecución de las $2k$ tareas seleccionadas con los dos procesadores.) Como k es un número fijo, la función $2kn+2^{2k}$ no es función exponencial, sino polinómica, de n, y crece moderadamente con relación a n. Por ejemplo, tomando $k=3$ resulta que $\omega_3/\omega_{opt} \leq 9/8$ y el trabajo de cómputo necesario es proporcional, a lo sumo, a $6n+64$. En general, dedicándole suficiente trabajo puede obtenerse el grado de precisión que se desee. Desde luego, la cantidad de trabajo necesaria crece rápidamente. Por ejemplo, para obtener un valor garantizado de ω_k que no supere al óptimo en más de un 2 por ciento puede ser necesario un tiempo de cómputo proporcional a $48n+2^{48}$, que ciertamente es suficiente para agotar más de uno –y de dos– presupuestos de gastos de computación.

Tal comportamiento no debería resultar muy sorprendente. Dado que para determinar soluciones óptimas parecen ser precisas cantidades de tiempo que crecen exponencialmente, es razonable que el costo de las soluciones aproximadas se comporte de forma semejante al ir aumentando la precisión garantizada de las soluciones. Lo sorprendente es que pueda evitarse el aumento exponencial de los tiempos de cómputo: hay un método para construir planificaciones de tareas independientes que garantiza resultados muy cercanos al óptimo y que requiere solamente tiempos de crecimiento polinómico. Oscar H. Ibarra y Chul Kim, de las universidades de Minnesota y Maryland, respectivamente, han puesto a punto hace poco un algoritmo eficiente para construir planificaciones con dos procesadores que tienen tiempos de terminación ω_k tales que $\omega_k/\omega_{opt} \leq 1 + 1/k$. La preparación del plan a partir del algoritmo requiere una cantidad de tiempo proporcional a $n + k^4 \log n$. (Cuando k y n son grandes, el valor de $n + k^4 \log n$



EN EL ALGORITMO DE PRIMER HUECO LIBRE se van alojando los pesos en las cajas $B_1, B_2,...$ en el mismo orden con que los pesos figuran en la lista de prioridad. (Si no hay lista de prioridad, los pesos se disponen arbitrariamente en lista.) Cada peso se coloca en la primera caja con capacidad para albergarlo. Dicho de otra forma, el peso se embala en la caja B_i de mínimo subíndice i tal que al sumarle el nuevo peso a los ya contenidos en la caja, el peso total no rebase el tope de carga prefijado. El empaquetamiento de la lista L que se muestra en la parte superior de la ilustración es notablemente poco eficiente: $FF(L)$, que es el número de envases necesarios, es igual a 17. En la parte inferior se muestra un envasado más eficiente. Este segundo envasado con 10 cajas es evidentemente óptimo, pues con él no se desperdicia espacio en ningún envase. El ejemplo exhibe también el peor rendimiento posible del algoritmo de primer hueco, porque en él se alcanza la cota para este algoritmo: $FF(L) \leq (17/10)OPT(L)$ siempre que para la lista L el número $OPT(L)$ sea múltiplo de 10. Los números de los múltiplos de pesos y las cajas van en color.



EL ALGORITMO DE PRIMER HUECO decreciente representa un avance sobre el algoritmo de primer hueco libre; en él, los pesos mayores son los primeros en ser envasados. En el algoritmo de hueco decreciente la lista de pesos a embalar se dispone en orden de pesos decrecientes, y después se aplica al algoritmo de primer hueco a la lista así reordenada. En el empaquetado producido por este algoritmo en la lista *L (arriba)* el número de cajas necesarias *FFD(L)*=7 es evidentemente óptimo, pues cada envase está lleno al máximo. Lo mismo que los demás algoritmos de planificación, los algoritmos de embalaje se encuentran expuestos también a comportamientos que resultan imprevisibles. Así, por ejemplo, si se retira un peso de la lista *(L) (abajo)*, resulta necesario utilizar una caja más para embalar por lista decreciente una lista más reducida.

suele ser mucho menor que 2^{2k}.) Sartaj K. Sahni, de la Universidad de Minnesota, ha generalizado el procedimiento para obtener algoritmos eficientes aplicables a más de dos procesadores. Tales procedimientos hacen uso de hábiles combinaciones de técnicas que transcienden de los objetivos de la presente exposición, pero este tipo de aproximación puede muy bien ser capaz de garantizar resultados cercanos a los óptimos en tiempos de cómputo razonables.

A pesar de que la investigación sobre la *NP*-completitud indica que no será posible, en general, hallar técnicas eficientes de construcción de planes óptimos de trabajo, existen muchos casos particulares interesantes de problemas de planificación que no son *NP*-completos y para los que es posible construir planificaciones óptimas en tiempos de crecimiento polinómico. Gran parte de las complicaciones que presentan los problemas de planificación son consecuencia de la intrincada estructura de las restricciones de precedencia y de las complicadas relaciones mutuas de los tiempos de ejecución. Limitando uno, o ambos, de tales factores pueden resultar tipos de casos particulares para los que pueden determinarse eficientemente planificaciones óptimas.

Por ejemplo, supongamos que se tenga un caso de planificación con un número arbitrario de procesadores, en el que todos los tiempos de procesamiento sean idénticos y el sistema de restricciones de precedencia sea arborescente, es decir, que cada tarea tenga a lo sumo una inmediatamente sucesiva. En este caso, el método de planificación por "caminos críticos", que es uno de los más utilizados, creará siempre planificaciones óptimas *(véase la ilustración superior de la página 79)*. En la planificación por caminos críticos la asignación de tareas a los procesadores se hace de acuerdo con la longitud de las diversas cadenas de precedencia encabezadas por dichas tareas en el diagrama de restricciones de precedencia. Las cadenas más largas de la parte no ejecutada del diagrama son las que tienen mayores sumas de tiempos de ejecución pendientes; se llaman por ello caminos críticos, porque sus tareas son las que probablemente provocarán embotellamientos en la ejecución del conjunto de tareas. En la planificación por caminos críticos se elige siempre como próxima tarea a realizar la que encabece el camino crítico en esa etapa de ejecución del trabajo.

T. C. Hu, de la Universidad de California en San Diego, demostró, en 1961, que la planificación por caminos críticos es óptima en el caso particular de restricciones de precedencia arborescentes y tiempos de ejecución idénticos. El resultado de Hu fue uno de los primeros en teoría de planificación. Debido a que en este caso particular todos los tiempos de ejecución son iguales, la planificación por caminos críticos consiste en asignar a cada tarea *T* un "nivel" *L(T)* igual al número de tareas de la cadena más larga encabezada por *T* *(véase la ilustración inferior de la página 79)*. En cuanto un procesador queda libre se le asigna la tarea de nivel más alto disponible en ese momento.

En otro caso particular de planificación no se impone ningún tipo de condiciones sobre la estructura de las restricciones de precedencia, pero ahora todos los tiempos de ejecución deben ser iguales y solamente debe haber dos procesadores para ejecutar el conjunto de tareas. En esta situación se dispone de varios métodos para determinar planes de trabajo óptimos. Uno de ellos, que en ocasiones se llama algoritmo *CG*, fue desarrollado en 1972 por Edward G. Coffman, de la Universidad de California en Santa Bárbara, y el autor. (*CG* denota Coffman y Graham.) Su inspira-

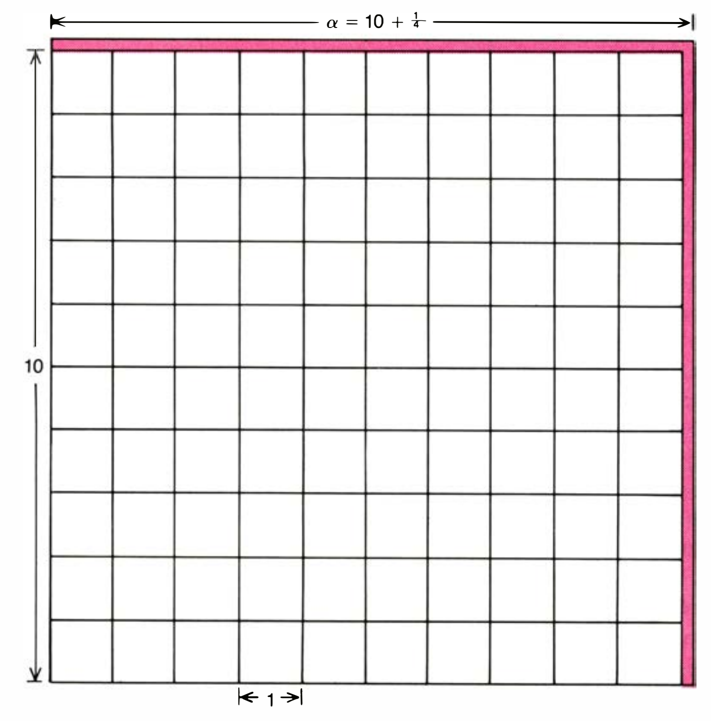
ción es semejante a la del algoritmo de niveles para el caso de restricciones arborescentes. Sin embargo, en el algoritmo *CG* el orden de ejecución de las tareas depende de la totalidad de cadenas encabezadas por cada una de ellas y no solamente de una cadena, la más larga, como ocurría en el algoritmo de niveles.

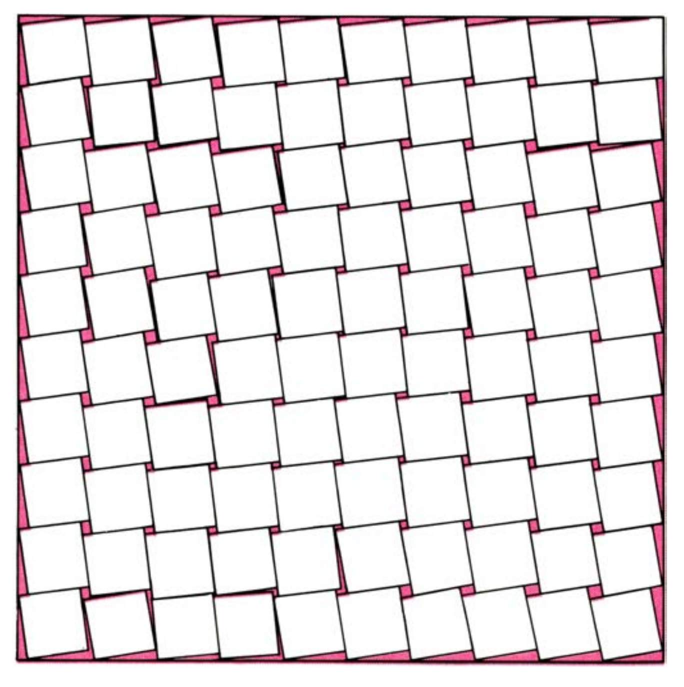
Antes de poder aplicar el algoritmo *CG* es preciso eliminar del diagrama de tareas las restricciones de precedencia superfluas. Por ejemplo, si $A \to B$ y $B \to C$, puede eliminarse la restricción $A \to C$. Para un conjunto de n tareas, este proceso de depuración de redundancias puede concluirse en menos de $n^{2,81}$ operaciones. El algoritmo *CG* empieza por numerar las tareas del conjunto *(véase la ilustración de la página 80)*. En primer lugar se le asigna el número 1 a una tarea que carezca de sucesoras. A continuación, para cada tarea con la propiedad de que todas sus sucesoras hayan sido numeradas ya, se forma una sucesión decreciente con los números asignados a todas sus sucesoras. La tarea cuya sucesión ocupe el primer lugar por orden "lexicográfico" entre las sucesiones construidas hasta ese momento será la próxima en recibir número. (En el orden "lexicográfico", la sucesión 5, 3, 2 va antes que las 6, 1; 5, 4, 2; 5, 3, 2, 1, etcétera.) Numeradas todas las tareas, de 1 a n, se construye la lista de prioridad situando las tareas por orden numérico decreciente. Se ha podido demostrar que la planificación construida con esta lista es siempre óptima para el caso particular de que haya dos procesadores y todos los tiempos de ejecución sean iguales. La razón fundamental que justifica el buen funcionamiento del algoritmo *CG* es que les asigna números tanto mayores, y, por consiguiente, tanta mayor prioridad, a las tareas que o bien tienen largas cadenas o bien tienen muchas sucesoras. La operación de numeración de las tareas puede efectuarse en aproximadamente n^2 operaciones, por lo que la eficiencia del algoritmo es muy alta.

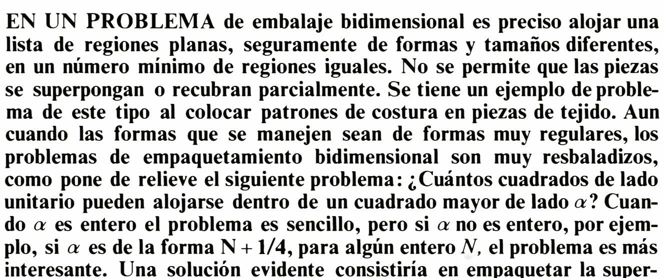
El algoritmo *CG* es solamente una de las diversas técnicas aplicables a problemas especiales con resultados óptimos. Quizá sea posible generalizar tales técnicas para obtener algoritmos igualmente fructíferos en problemas particulares, como, por ejemplo, el problema de tres procesadores con un conjunto de tareas del mismo tiempo de ejecución. Es preciso observar, sin embargo, que incluso el caso ligeramente menos particular de que haya dos procesadores con un conjunto de tareas cuyos tiempos de ejecución sean, o bien de una unidad, o bien de dos unidades, es *NP*-completo, como ha sido recientemente demostrado.

Solamente he analizado hasta ahora un tipo de problema de planificación, pero los problemas de planificación se presentan en muchos lugares y bajo muchas formas diferentes. Uno de los problemas más interesantes se enuncia volviendo del revés el modelo fundamental que acabamos de estudiar: en lugar de mantener fijo el número de procesadores e intentar que el tiempo de terminación sea mínimo, se trata ahora de completar la realización de un conjunto de tareas en un plazo fijo, utilizando el menor número posible de procesadores. Dicho de otra forma, el problema pide con cuántos procesadores podremos arreglarnos para ejecutar un conjunto dado de tareas antes del vencimiento de un plazo prefijado.

Cuando las tareas son independientes este problema de planificación se enuncia de forma distinta, y suele denominarse "problema de embalaje". En un modelo del problema de embalaje hay un conjunto de artículos $I_1, \ldots, I_r$; cada







EN UN PROBLEMA de embalaje bidimensional es preciso alojar una lista de regiones planas, seguramente de formas y tamaños diferentes, en un número mínimo de regiones iguales. No se permite que las piezas se superpongan o recubran parcialmente. Se tiene un ejemplo de problema de este tipo al colocar patrones de costura en piezas de tejido. Aun cuando las formas que se manejen sean de formas muy regulares, los problemas de empaquetamiento bidimensional son muy resbaladizos, como pone de relieve el siguiente problema: ¿Cuántos cuadrados de lado unitario pueden alojarse dentro de un cuadrado mayor de lado α? Cuando α es entero el problema es sencillo, pero si α no es entero, por ejemplo, si α es de la forma N + 1/4, para algún entero N, el problema es más interesante. Una solución evidente consistiría en empaquetar la superficie de $\alpha \times \alpha$ rellenando un cuadrado de $N \times N$ con N^2 cuadrados, y sacrificando el área no cubierta *(en color)*, que mide casi $\alpha/2$ unidades cuadradas como desecho inevitable. Tras experimentar otros métodos de empaquetamiento *(derecha)* sentimos la tentación de concluir que no puede conseguirse mejora alguna sobra la solución trivial, pero sorprendentemente no es así. Paul Erdös, Hugh Montgomery y el autor han demostrado recientemente que cuando α crece suficientemente, existen empaquetamientos de un cuadrado de $\alpha \times \alpha$ que no dejan más de $\alpha^{0,634\ldots}$ unidades cuadradas de área descubierta, sensiblemente menos que las $\alpha/2$ unidades despilfarradas con el recubrimiento trivial. No se ha determinado aún el área mínima que permanecerá descubierta cuando α sea muy grande, aunque probablemente $\alpha^{0,5}$ sea una conjetura verosímil.

artículo I_k tiene un peso w_k. El problema consiste en embalar todos los artículos en un número mínimo de cajas B_1, B_2,... de forma que el peso total de los artículos contenidos en cada caja no supere un cierto peso fijo W. (Con la terminología del modelo fundamental de planificación, los artículos son tareas, los pesos son sus tiempos de ejecución, las cajas son procesadores y el peso fijo máximo representa el plazo de terminación, fijo.)

El problema de embalaje surge, bajo diversidad de formas, en muchos casos prácticos. Por ejemplo, un fontanero ha de cortar un conjunto de tubos de longitudes diferentes utilizando el mínimo número posible de tubos de longitud normalizada; una cadena de televisión desea pasar los "comerciales" contratados, cuya longitud es variable, interrumpiendo el menor número posible de veces su programa; un fabricante de papel debe proveer a sus clientes con rollos de papel de distintos anchos, que se cortan del mínimo número posible de rollos de dimensión normalizada. En general, los problemas de embalaje son extraordinariamente difíciles. Por el momento, los únicos métodos conocidos para la obtención de embalajes óptimos (que son los que precisan del mínimo número de cajas) consisten, en esencia, en examinar todas las posibles formas de empaquetar los artículos y elegir después la mejor. Al igual que la mayoría de los problemas de planificación, el problema de embalaje es *NP*-completo, y por ello es de temer que todos los algoritmos generales destinados a obtener soluciones óptimas sean de crecimiento exponencial. Así pues, muchos de los algoritmos creados para este tipo de problemas se han diseñado con la intención de obtener métodos de empaquetado razonablemente aproximados al óptimo.

Al estudiar el problema de embalaje es conveniente ordenar en una lista, situándolos arbitrariamente en ella, los pesos de los artículos: $L=(w_1, w_2,..., w_r)$. Como no hay restricciones de precedencia no hay peligro de confusión al representar los artículos mediante sus pesos, por lo que puede considerarse que L representa una lista de los artículos a embalar. Un método evidente de empaquetar los pesos de L es el llamado algoritmo de "primer hueco libre" (first-fit algorithm) *(véase la ilustración de la página 82)*. Con las reglas de esta algoritmo, se van colocando los pesos en cajas, en el mismo orden con que figuran en L: primero w_1, luego w_2, y así sucesivamente. Cuando le llega a w_k el turno de ser embalado, se le asigna la primera caja que pueda contenerlo, es decir, la caja B_i de mínimo índice i capaz de alojar el peso. (Un peso w_k puede alojarse en una caja si al sumar el peso w_k a los ya colocados en ella la suma total no supera W.)

¿Hasta qué punto es eficiente este algoritmo? Dicho de otra forma, si se representa mediante $FF(L)$ el número de cajas necesarias al aplicar a L el algoritmo de primer hueco libre, y se denota $OPT(L)$ el número de envases necesarios en un empaquetado óptimo de los pesos de L, ¿hasta qué punto puede ser $FF(L)$ mayor que $OPT(L)$? En 1973, Jeffrey D. Ullman, de la Universidad de Princeton, descubrió que, cualquiera que sea la lista L, $FF(L) \leq (17/10)OPT(L)+2$. Ullman demostró también que el coeficiente 17/10 no puede ser mejorado. No obstante, cuando $OPT(L)$ sea múltiplo de 10, la constante 2 puede suprimirse en la desigualdad: $FF(L) \leq (17/10)OPT(L)$. Se ha enunciado la conjetura de que esta cota simplificada es válida en todos los casos.

La cota superior que acabamos de exponer muestra que el algoritmo de primer hueco puede tener un rendimiento bastante bajo: puede llegar a requerir hasta un 70 por ciento más del número óptimo de envases necesarios. Experimentando con este algoritmo se aprecia que los peores resultados se obtienen colocando grandes pesos al final de la lista, obligando a empezar nuevas cajas, aún quedando grandes disponibilidades de espacio en cajas parcialmente rellenas. Parece razonable reordenar la lista, colocando al principio todos los pesos grandes, con el fin de que los pesos pequeños situados al final sirvan para rellenar los huecos más o menos aleatorios de las cajas casi llenas. Esta idea sugiere un nuevo procedimiento de empaquetado, llamado algoritmo de primer hueco libre decreciente. La lista L presenta los pesos ordenados en sucesión decreciente, aplicándose después el algoritmo de primer hueco libre. Este algoritmo es francamente bueno *(véase la ilustración de la página 83)*. Si $FFD(L)$ representa el número de cajas precisas para empaquetar L según el nuevo algoritmo de lista decreciente, puede demostrarse que $FFD(L) \leq (11/9)\ OPT(L)+4$, cualquiera que sea la lista L. Se ha demostrado que es imposible mejorar el coeficiente 11/9.

La expresión $(11/9)OPT(L)+4$ tiene un aspecto engañosamente sencilla. Se han encontrado serias dificultades para demostrar que dicha expresión es efectivamente cota superior de $FFD(L)$. Por el momento, la única demostración que se conoce es la preparada por David S. Johnson, de los Bell Laboratories.

Cuando un problema de embalaje requiere gran número de cajas, la constante 4 de la desigualdad resulta relativamente insignificante. En tal caso, el algoritmo de primer hueco decreciente garantiza el empaquetado de cualquier lista en no más del 22 por ciento de exceso sobre el número óptimo de cajas. Este resultado es ciertamente muy superior al aumento del 70 por ciento sobre el óptimo que puede presentarse con el empaquetado por "primer hueco" de una lista particularmente difícil de manejar.

Aunque obligados a permanecer dentro de sus respectivas cotas establecidas, los algoritmos de "primer hueco" o de "primer hueco decreciente" están expuestos, lo mismo que otros procedimientos de planificación, a comportamientos imprevisibles cuando se alteran los parámetros del modelo. Por ejemplo, puede ocurrir que con el algoritmo de primer hueco aumente el número de envases necesarios al suprimir uno de los pesos. Si se denota L' la lista así reducida, se desconoce todavía cuánto puede valer el cociente $FFD(L')/FFD(L)$, y se desconoce también si $FFD(L)$ puede aumentar en algún caso al suprimir el máximo elemento de L. Como antes ocurría, la causa del imprevisible comportamiento de los algoritmos de embalaje hay que buscarla en la condición de que cada peso se vaya alojando en el primer envase disponible, pero tal tipo de comportamiento no es nada infrecuente en las situaciones, mucho más complejas, de planificación en el mundo real.

El sencillo modelo de planificación que acabo de describir ha proporcionado gran cantidad de información relativa a problemas cuya complejidad remeda la de situaciones reales. Son posibles muchas generalizaciones del modelo fundamental. Se puede alterar el modelo para permitir la interrupción de tareas antes de terminación o para permitir que haya operadores inactivos aun existiendo tareas disponibles para ellos.

Puede tomar en consideración otros recursos, aparte de la capacidad de procesamiento, necesarios para la ejecución de las tareas, así como tiempos de ejecución aleatorios, descritos en términos de probabilidad, diversas medidas de rendimiento del modelo, etcétera. Sometiendo estos modelos más generales al mismo tipo de análisis que he explicado en estas páginas, los investigadores están hoy avanzando rápidamente en la comprensión de los difíciles problemas de planificación.

Bruegel el Viejo, guía para el estudio de la ciencia del s. XVI

El gran artista flamenco sentía un profundo interés por las ideas científicas y por las máquinas de su tiempo. Consecuencia de esto es que muchas de sus obras ofrecen atisbos acerca de los conocimientos prácticos de hace 400 años

H. Arthur Klein

Pieter Bruegel el Viejo, el gran artista flamenco del siglo XVI, es famoso, entre otras cosas, por la abundancia y claridad de los detalles de sus pinturas y dibujos para grabados sobre temas alegóricos. Esta característica de su obra ofrece un atisbo poco corriente de la ciencia y la tecnología del siglo XVI. Pueden encontrarse datos de este tipo en algunas de sus pinturas (entre las cincuenta, aproximadamente, que conservamos de este artista), en sus dibujos (existen casi ochenta) y en varios de sus ciento setenta grabados en lámina de cobre, realizados siguiendo la pauta de dibujos hechos por el propio Bruegel. Es preciso, no obstante, examinar cuidadosamente su obra si queremos encontrar la información aludida.

Dos pinturas realizadas por Bruegel en 1563 muestran técnicas del siglo XVI para la construcción en gran escala. Ambas se titulan *La Torre de Babel.* La "pequeña *Babel*" constituye, hoy, el tesoro principal del Museo Boymans-van Beuningen de Rotterdam. La "gran *Babel*" es una de las catorce obras maestras de Bruegel que se conservan en el Kunsthistorisches Museum de Viena.

Bruegel empezó a trabajar en este tema bíblico diez años, por lo menos, antes de terminar sus pinturas. El *Génesis* refiere que, después del Diluvio, los descendientes de Noé empezaron a construir una torre, en la tierra de Sinar, cuya cumbre debía llegar "al cielo". Pero Dios terminó con este esfuerzo impío haciendo que los constructores hablaran lenguas tan distintas que les fuera imposible comunicarse entre sí. Fueron dispersados por toda la tierra y la torre quedó inacabada.

Bruegel utilizó, al menos, tres fuentes, además de la Biblia, para representar este relato justificativo de la multiplicidad de lenguas habladas por la humanidad. La primera es el Coliseo de Roma, que vio durante una visita a Italia en su juventud. La segunda son los escritos de Herodoto. Por último, Bruegel incorporó un cierto número de máquinas y métodos utilizados en su tiempo en los principales puertos de los Países Bajos.

Las pinturas sobre este tema se enfrentan de modo atrevido con el problema de cómo se pudo planear un edificio tan soberbio. Lo representan enorme pero incompleto. Sus partes terminadas sobrepasan ya las nubes. Su estructura interna aparece como una serie de arcos de diseño muy semejante a los del Coliseo. Este antiguo anfiteatro había sido construido formando una intrincada trabazón de más de 1000 arcos de piedra. 240 de éstos podían verse desde el exterior, dispuestos en tres grandes hileras elípticas. Otros 240 grandes arcos se encontraban situados radialmente en torno a elipses interiores. Estas y las elipses exteriores estaban unidas entre sí sólidamente por un sistema de 480 arcos radiales.

La anatomía interna del Coliseo fue adaptada por Bruegel para justificar la altura enhiesta y la rigidez de su Babel imaginaria. Pero sus dos representaciones de Babel muestran una amplia rampa espiral que se enrosca en torno a la torre en construcción, rampa que no existía en el Coliseo. No obstante, según Herodoto, una rampa similar serpenteaba en torno al exterior del gran zigurat, o torre sagrada, babilónico, denominado Etemenanki. Es muy difícil que Herodoto hubiera visto en persona el zigurat, pero basó su descripción en informaciones obtenidas de gentes que afirmaban haberlo visto: "una torre de mampostería sólida sobre la que se alzaba una segunda torre y, sobre ésta, una tercera y así sucesivamente hasta llegar a la octava". Según su relato, el camino que llevaba desde la base hasta la cúspide tenía la forma de una rampa espiral que "serpenteaba en torno a todas las torres".

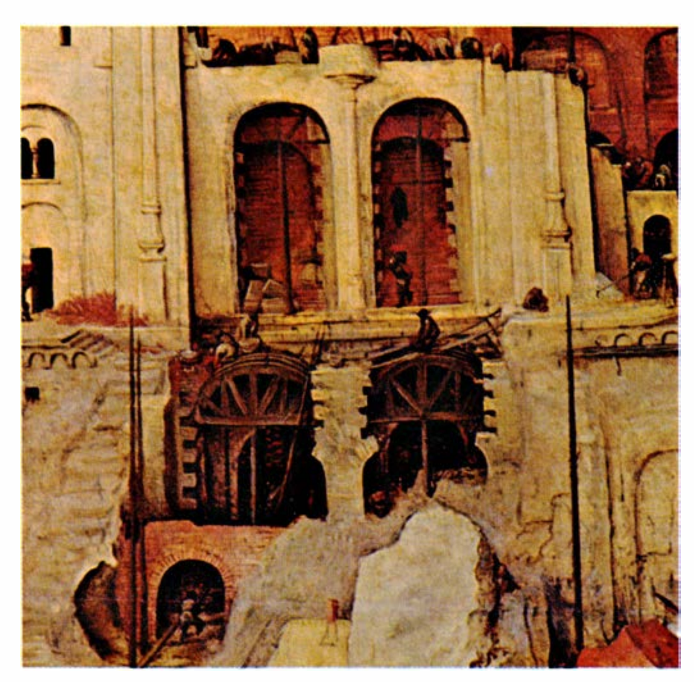
Las pinturas de Bruegel en las que se representa la torre de Babel muestran (la de Rotterdam de forma muy clara, la de Viena en menor grado) una enorme rampa similar que sube en espiral en torno a la maciza torre. El hecho de que ésta hubiera descrito ocho vueltas en el caso de haberse terminado la construcción del edificio, sugiere la existencia de una relación entre Herodoto y Bruegel.

El aspecto más significativo de estas pinturas, en el contexto que estamos analizando, radica en el recurso de Bruegel a entrelazar técnicas y artificios propios de su ambiente flamenco-holandés. Las dos torres gigantescas se alzan junto a un puerto activo, con lo que se logra responder a la pregunta de cómo pudieron reunirse todos los materiales de construcción. Los puertos se encuentran junto a un gran brazo de agua o estuario, tal como sucedía en el puerto de Antwerp, activo y en plena expansión, la ciudad en la que Bruegel vivió durante cierto número de años.

En el muelle de la pequeña *Babel* se distingue, entre el amasijo de mercancías y materiales, la masa tosca de una gran máquina que se inclina hacia arriba y hacia un lado, más allá de su rueda lateral, y parece, a primera vista, un antiguo y monstruoso cañón de asedio. Si se examina más de cerca, se observa que se trata de una inmensa grúa de madera o cabria. La gran rueda lateral es un mo-

"TORRE DE BABEL", pintada al óleo por Pieter Bruegel el Viejo sobre una tabla de madera de roble, representa el relato bíblico de Babel referido en el libro del Génesis y muestra también una serie de técnicas de construcción del siglo XVI. Estas incluyen los mecanismos elevadores movidos por tracción humana que se encuentran en el lado derecho de la torre, en los niveles primero y segundo de la rampa. Este cuadro, que se encuentra en el Kunsthistorisches Museum de Viena, es conocido como "la gran *Babel*" para distinguirlo de otro similar, "la pequeña *Babel*".





LOS DETALLES DEL CUADRO "BABEL" incluyen un mecanismo elevador en forma de tambor *(izquierda)* que se encuentra en el primer nivel de la rampa, por debajo de la grúa reproducida en la portada de este número, así como dos arcos *(derecha)* en construcción cerca del centro del cuadro. El mecanismo elevador, igual que la grúa, era movido por hombres situados en el interior de un molino de rueda de escalones y se basaba en un sistema de poleas. Tenía menor capacidad elevadora que la grúa. Para esta obra, Bruegel se inspiró en la tecnología de su tiempo.

lino de rueda de escalones giratorio, con una disposición semejante a la de una jaula de ardillas en la que trabajadores humanos proporcionaban la energía necesaria para operar la grúa. Bruegel indicaba de esta manera cómo se elevaban hasta su destino las piedras de la alta torre.

La autenticidad del aparato como un artificio propio del siglo XVI para elevar pesos está documentada, además, por una pintura de otro artista de la época, Pieter Pourbus el Viejo. El cuadro es un retrato de Jan Fernaguut, un mercader de Brujas. Por encima del hombro del mercader, como si se tratara de una ventana abierta, se distingue una parte del muelle de carga cuyo comercio ayudó a enriquecerle. Alzándose frente a los edificios del muelle se encuentra el mismo tipo de estructura de madera inclinada con forma similar a la de alguna extraña clase de ave picuda. La ancha rueda de madera situada a un lado, aparentemente de casi seis metros de diámetro, está cubierta parcialmente por un tejado inclinado que permitiría trabajar a los obreros del molino de rueda de escalones incluso con mal tiempo.

En la gran *Babel,* Bruegel situó una inmensa grúa de este tipo, movida con energía humana, en el segundo piso de la rampa espiral. La grúa tiene un molino de rueda de escalones a cada lado. Si se estudian de cerca los pequeños detalles se constata que por lo menos seis y posiblemente ocho obreros trabajan en el interior de las ruedas. La grúa del cuadro está elevando una gran losa de piedra.

¿Podían levantarse con grúas de este tipo todas las pesadas piedras de la torre? Un obrero varón, adulto y sano, puede suministrar una potencia de poco más de un caballo de vapor, o, aproximadamente, 75 vatios, durante un largo período de tiempo del orden de varias horas. No obstante, en esfuerzos más breves, puede llegar a producir hasta tres caballos, o 225 vatios. Si seis obreros estuvieran trabajando a máximo rendimiento en el interior de un molino de rueda de escalones, podrían desarrollar entre 1,5 y 2 caballos de vapor durante un lapso de tiempo limitado: algo así como unos pocos minutos seguidos por un período de descanso.

Incluso si admitimos unas pérdidas del orden de un 20 por ciento en fricción producida en los soportes y en los cordajes de las poleas de las grúas, nos queda una red comprendida entre 1,2 y 1,6 caballos de vapor. Dado que un caballo de vapor equivale a 76 kilográmetros por segundo, 1,2 caballos serán 91,2 kilográmetros por segundo. Para elevar una losa de piedra de una tonelada unos quince metros se requieren 13.825 kilográmetros de trabajo o energía. Bajo las condiciones antes descritas, se podría alcanzar la cantidad mencionada en 151,5 segundos, o dos minutos y medio, aproximadamente.

Los dos magníficos cuadros que representan la torre de Babel muestran, con bastante claridad, media docena, por lo menos, de máquinas elevadoras movidas por tracción humana; algunas más son sugeridas de forma menos distinta. Uno de los artificios más claramente dibujados se alza en la rampa que se encuentra por debajo de la grúa. Se trata simplemente de un enorme tambor en torno al cual se enrollan unas cuerdas a medida que unos hombres lo hacen girar. Con la ayuda de poleas, este artificio podía elevar objetos algo menos pesados que aquellos que caen dentro de la competencia de las grúas.

En 1556, Bruegel empezó la serie de dibujos que hoy en día son ampliamente conocidos como Los Siete Pecados Capitales. *Avaricia* fue grabada y publicada en 1556. En 1557 se publicaron los grabados de sus dibujos *Soberbia, Envidia, Ira, Gula, Lujuria* y *Pereza.* El gran éxito obtenido por los Pecados llevó ineludiblemente al artista a realizar una



"LA PARABOLA DE LOS CIEGOS", pintada en 1568, ilustra un relato neotestamentario. Revela la atención que Bruegel prestaba a los detalles, ya que los médicos han diagnosticado cinco tipos distintos de ceguera a partir de su representación en los cinco rostros visibles aquí.

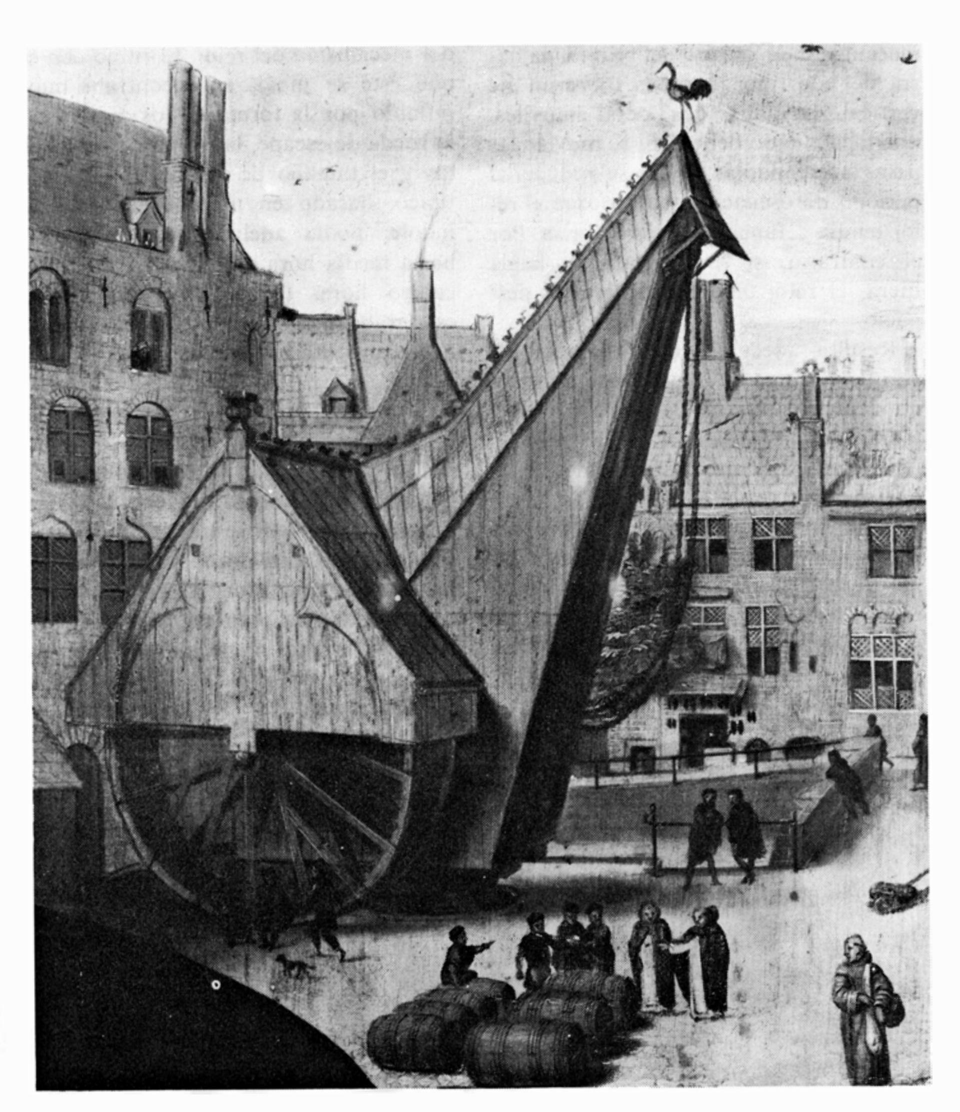
serie de las Siete Virtudes. Estos dibujos, mucho más sobrios y sombríos, fueron realizados por Bruegel en 1559 y 1560. Las virtudes representadas en ellos son fe, esperanza, caridad, justicia, prudencia, fortaleza y templanza.

El grabado *Avaricia* revela lo que podría denominarse "el curioso caso de las flechas de ballesta". En una parte del grabado, varios pecadores dan muestras de su avaricia disparando con ballestas a una enorme bolsa colgada tentadoramente por encima de sus cabezas. Las flechas ascienden sin trazar curva alguna para luego dar la vuelta y caer perpendicularmente.

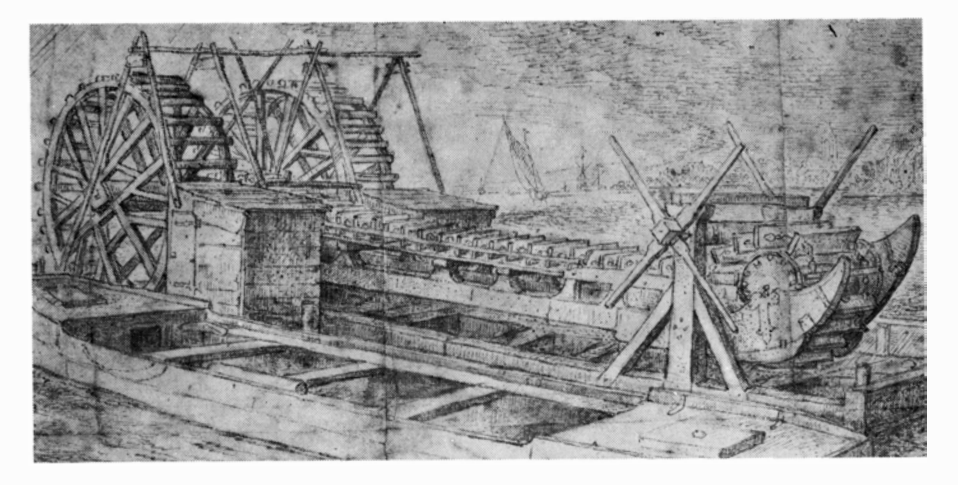
Bruegel, un artista extraordinariamente fiel en todos sus detalles físicos y naturales, ha indicado aquí unas trayectorias distintas de las que serían demostradas más tarde por la ciencia balística. El curso que siguen las flechas en su vuelo se muestra de acuerdo, no obstante, con las ideas erróneas predominantes en su tiempo. Bruegel dibujaba su representación del tema mucho antes de que Galileo descubriera y publicara que las balas de cañón, piedras y flechas, cuando son proyectadas, describen trayectorias parabólicas, al menos en la medida en que se lo permite la resistencia del aire.

En el ciclo sobre los Pecados, Bruegel abordó no sólo problemas relativos al movimiento sino también aspectos del tiempo. El tema aparece de modo específico en el grabado *Pereza*. Allí se ve a la Reina Pereza que yace en actitud embrutecida y rodeada de caracoles, babosas y otros símbolos adecuados al tema.

En la parte superior derecha se encuentra un gigantesco aparato de relojería colgado de un árbol. Una figura humana, o humanoide, situada frente al reloj, golpea una campana con un martillo (despertador que nos recuerda a nosotros y a los que cometen el pecado de pereza que es más tarde de lo que pensamos). Pueden verse en este extraño reloj varios mecanismos grandes. Son visibles también las pesas gemelas que suministraban la fuerza motriz a este tipo de relojes en tiempo de Bruegel. Otro elemento esencial, que aparece justo por debajo del pie del hombre que toca la campana, es una barra de hierro, parcialmente escondida, de la que cuelga una pesa metálica que puede colocarse en varias posiciones gracias a la existencia de dientes en la barra. Pese a su situación imposible en el árbol, puede reconocerse en este mecanismo un foliot, una barra oscilante que controlaba el



GRUA MOVIDA POR TRACCION HUMANA pintada en el siglo XVI por el artista flamenco Pieter Pourbus el Viejo (1523-1584); muestra con algo más de detalle el tipo de mecanismo elevador representado por Bruegel en la gran *Babel*. Ocupa una esquina del retrato de Jan Fernaguut de Brujas y trata de mostrar un aspecto del comercio portuario. La linea curva y obscura en el extremo inferior izquierdo es el hombro de Fernaguut. El molino de rueda de escalones (probablemente había un segundo molino al otro lado) está cubierto por una techumbre para cobijar a los hombres que suministraban la energía necesaria para mover la grúa. En la base de ésta se ve un pivote gracias al cual toda la estructura podía hacerse bascular hacia un barco situado en el muelle adyacente.



DRAGA MOVIDA POR TRACCION HUMANA pintada por el artista flamenco Roeland Savery (1576-1639); da muestras de la precisión con la que Bruegel representaba los mecanismos elevadores del siglo XVI. Los dos molinos de rueda de escalones situados a la izquierda movian una correa de transmisión con cucharones de draga que penetraban en el agua. Este dibujo, de la Trustees of the Chatsworth Settlement, está en la Colección Devonshire en Chatsworth (Inglaterra).

mecanismo de escape del reloj. Una barra de este tipo, montada sobre un eje vertical, sujetaba dos pesas móviles, una a cada lado del eje. Si se movían las pesas acercándolas al eje se reducía el período de oscilación, con lo que el reloj tendía a funcionar más de prisa. Por el contrario, si se desplazaban hacia fuera, el reloj tendía a andar más despacio.

Resulta adecuado decir "tendía a", ya que el ritmo de movimiento del reloj no dependía sólo de la posición o del tamaño de las pesas. El eje de la barra llevaba paletas o proyecciones metálicas, que detenían y ponían en marcha, alternativamente, una rueda dentada, denominada habitualmente rueda de escape. Bruegel nos la representa como una rueda con dientes en forma de anzuelo que aparece en el árbol entre las partes del mecanismo del reloj. El ritmo con el que éste se movía se encontraba muy influido por la forma de los dientes de la rueda de escape, la forma de las paletas y el tamaño de las pesas. Un reloj típico, basado en mecanismos de esta índole, podía adelantarse o atrasarse hasta media hora en un plazo de veinticuatro horas (cuando Bruegel dibujó este grabado aún no se conocía la mejora que representó el péndulo).

El carácter primitivo de este reloj se ve subrayado por la monstruosa y simbólica esfera situada en la parte superior izquierda del grabado. Tiene sólo una manecilla (en rigor se trata de una mano situada al extremo de un brazo humano). Los relojes que Bruegel conocía mejor tenían sólo una manecilla horaria. Los relojes con dos manecillas existían ya en su tiempo, pero ni abundaban ni eran económicos. Una manecilla que indicara los minutos tenía escasa utilidad cuando podía esperarse que el mecanismo funcionara muy de prisa o muy despacio en un período de tiempo de veinticuatro horas.

Otros símbolos que combinan características mecánicas y humanas aparecen en el grabado dibujado por Bruegel en torno a la *Gula*. Una cabeza inmensa y horrible es, al mismo tiempo, un molino de viento. Unos obreros minúsculos arrastran sacos de grano y de otros materiales hasta sus mandíbulas abiertas donde serán molidos (o sea masticados y tragados). Queda de manifiesto el apremio a devorar mecánicamente, incluso sin necesidad.

En el grabado *Lujuria* encontramos un mecanismo relativamente sofisticado:



EL PECADO DE PEREZA fue dibujado por Bruegel en 1557 formando parte de una serie de dibujos para grabados sobre los siete pecados capitales. Varios de los mecanismos de un reloj aparecen en la parte superior derecha: entre ellos se encuentra el foliot situado bajo el pie del personaje que, en posición horizontal, toca la campana. En la parte superior izquierda aparece la esfera de un reloj con una sola manecilla (en realidad un solo brazo). La mayoría de los relojes del siglo XVI tenían sólo una manecilla horaria ya que no eran suficientemente precisos para que un minutero tuviera alguna utilidad. Cerca del centro del grabado se encuentra una noria movida por debajo por la corriente de agua. Probablemente Bruegel utilizó estos ingenios mecánicos, desarrollados ya en su época, para simbolizar que los perezosos están perdiendo el tiempo.

una fuente decorada y con dos chorros gemelos suministra agua a un estanque en el que chapotean y se abrazan unas parejas desnudas. La corriente de agua que serpentea en este paisaje de lujuria tiene su origen más allá de la fuente y hace girar una noria parcialmente sumergida en torno a la cual pueden verse a los esclavos de la lascivia.

En varios dibujos de Bruegel los demonios llevan a cabo guerras simbólicas. En *Fortaleza*, un dibujo perteneciente a la serie de las Virtudes, se encuentran en acción sangrienta unas huestes de soldados tanto a pie como a caballo. Pueden verse dos grandes escudos de combate fabricados en madera. Una de estas formas primitivas de armadura de madera lleva en su centro un martillo articulado dispuesto a aplastar al adversario que se acerque. Cerca del centro del dibujo una máquina de guerra similar ha sido derribada; su martillo pende, ya inútil, y está a punto de ser destrozado por la caballería que avanza.

El grabado (realizado probablemente por Philipp Galle) del dibujo de Bruegel sobre la *Templanza* parece revelar, de la forma más cumplida, el aprecio que el artista sentía por la tecnología, la ciencia y el arte de su tiempo. La Templanza, una figura alta y con aspecto regio, lleva sobre su cabeza un reloj, como si de una corona se tratara, el artilugio más sofisticado de su tiempo. Su pie derecho se apoya sobre el aspa de un molino de viento, la fuente de energía más avanzada. En la mano izquierda sujeta un par de anteojos que representan un instrumento óptico de primera importancia en una época en la que el telescopio y el microscopio compuesto pertenecían aún al futuro.

En torno a Templanza se encuentran grupos dedicados a actividades que pueden considerarse estimadas por el artista y por los clientes que compraban sus grabados. En la parte inferior derecha se enseña a leer en una escuela. En el extremo inferior izquierdo tienen su desarrollo el cálculo, la teneduría de libros y la contabilidad. Por detrás de esta escena, un artista se dedica a pintar. Más arriba, grupos instrumentales y corales se entregan a la música. En la parte superior derecha aparece un escenario donde se desarrolla un drama incluso con apuntador. Cinco hombres con aspecto serio y formal, probablemente teólogos, están enfrascados en una seria discusión en el centro a la derecha, por encima de la escuela. Representa la retórica, el arte de



LA VIRTUD DE LA TEMPLANZA fue dibujada por Bruegel para realizar este grabado. El dibujo incluye cierto número de temas relativos a la ciencia y a la tecnología. En el centro de la parte superior se simbolizan la astronomía y la geodesia. En la columna situada a la derecha un hombre deja caer una plomada y otro calcula distancias. Otros tipos de medidas de superficies se desarrollan en la zona situada por detrás de la columna. El instrumento consistente en una rueda situada al extremo de un bastón proporciona un medio para medir ángulos. La Templanza se encuentra asociada con algunos símbolos de la tecnología: incluyen un reloj, un par de anteojos y el aspa de un molino de viento. La favorable acogida que el público dio a los grabados sobre los siete pecados movió a Bruegel a realizar siete dibujos sobre las virtudes.



EL PECADO DE LUJURIA fue otro de los dibujos de Bruegel para sus grabados sobre los siete pecados capitales. Este detalle del grabado muestra un estanque alimentado por una fuente con dos chorros gemelos. Es uno de los varios lugares del dibujo en los que se patentiza la lujuria o lascivia (otro título que se da a veces al grabado). Posiblemente el agua era elevada hacia los caños mediante una noria movida por la corriente de agua. Una noria que, aparentemente, está asociada con un molino aparece también en el grabado. En el grabado *Avaricia,* que no aparece aqui, revela lo que podría denominarse "el curioso caso de las flechas de ballesta". En una parte del grabado, varios pecadores dan muestras de su avaricia disparando con ballestas a una enorme bolsa colgada tentadoramente por encima de sus cabezas. Las flechas ascienden rectas y caen perpendicularmente.



EL CONCEPTO DE FORTALEZA, uno de los temas representados por Bruegel en su serie de dibujos para grabados sobre las siete virtudes, fue expresado situando el símbolo alado de la fortaleza en medio de batallas y matanzas. Este proyecto dio a Bruegel la ocasión de mostrar una serie de artificios militares metálicos como este escudo móvil. Pueden verse dos grandes escudos de combate fabricados en madera. Una de estas formas primitivas de armadura de madera lleva en su centro un martillo articulado dispuesto a aplastar al adversario que se acerque. Cerca del centro del dibujo una máquina de guerra similar ha sido derribada; su martillo pende, ya inútil, y está a punto de ser destrozado por la caballería que avanza. Las virtudes dibujadas por Bruegel incluyen las virtudes teologales (fe, esperanza y caridad), así como las llamadas virtudes cardinales.

la presentación y de la persuasión eficaz a través del lenguaje.

En el extremo superior derecho se encuentran los símbolos de la balística y del estudio de las trayectorias. Un ballestero apunta a un blanco situado encima de un mástil. Lo mismo hace otro individuo con un primitivo mosquete muy largo. Por detrás de estos dos personajes aparecen las bocas de dos grandes cañones con ruedas; cerca de ellos se encuentran unos montones de balas de cañón.

Dos hombres toman medidas sobre una columna aislada. El que se encuentra encima de la misma deja caer una plomada para asegurarse de su verticalidad. El hombre que se encuentra sentado sobre una tabla en mitad del pilar utiliza un compás de puntas para determinar distancias. Muy por detrás de la columna, en el campo abierto, parece desarrollarse un proceso de medida de superficies. Algo a la izquierda de la columna, un hombre mira a lo largo de un bastón rematado por una rueda. Este artificio proporcionaba un medio simple, pero muy preciso, para medir ángulos a simple vista. Por detrás de este hombre se halla un grabador-pintor que utiliza una escuadra para verificar los ángulos de un cuadro.

Más allá de la cabeza de Templanza se encuentra un globo en el que se sugieren vagamente los continentes y océanos. De pie, a su derecha, el geógrafo-cartógrafo mide distancias en su superficie con ayuda de un compás de puntas. Por encima del globo aparece un astrónomo que empuña su compás tratando, según parece, de determinar la distancia hasta la luna.

Cuando andaba yo en busca de una buena portada para mi libro *The World of Measurements* ("El mundo de las medidas"), no encontré nada mejor que utilizar algunos fragmentos del grabado *Templanza.* Tres fragmentos me bastaron, ya que en ellos se compendiaban las medidas, unidades y conceptos gracias a los cuales ha avanzado la ciencia.

En varias de sus obras Bruegel mostró también que estaba atento a las principales dolencias que afligían a sus semejantes, así como a los métodos utilizados por médicos y curanderos para tratar las enfermedades tanto físicas como mentales. Un grabado realizado en 1559, a partir de su dibujo sobre *La Bruja de Mallaghem,* sugiere la locura colectiva de las gentes crédulas. Un "curandero" trabaja en medio de una muchedumbre caótica de flamencos frenéticos y extrae tumores en forma de piedra de la cabeza

de los enfermos. Bruegel exploró también la credulidad y la histeria populares en un cuadro titulado *Peregrinación de los epilépticos a la iglesia de Molenbeek* (es el nombre de una ciudad flamenca). Esta obra se conserva aún hoy en forma de dos grabados, realizados después de su muerte, y de un dibujo (probablemente también una copia póstuma) que se encuentra asimismo en la colección Albertina de Viena. En los grabados, unas mujeres, que parecen más histéricas que epilépticas, avanzan aguantadas y sujetas por unos hombres que las cuidan. Las llevan a un puente donde bailarán a los sones de una gaita hasta que suden o consuman sus enfermedades.

Otros dos cuadros de Bruegel tienen importantes implicaciones médicas. *Los Mendigos,* pintado en 1568, muestra cuatro figuras tullidas y contorsionadas, que se apoyan en muletas y bastones y que van extrañamente vestidas con ropas adornadas con colas de zorro. Algunos comentaristas han sugerido que estos hombres muestran varios síntomas y fases de lepra.

En *La Parábola de los Ciegos,* Bruegel representa a seis ciegos que se guían el uno al otro por razones de seguridad. El primero del grupo ha caído en una zanja y los otros le van a seguir con toda seguridad. Los médicos han diagnosticado un distinto tipo de ceguera en cada uno de los cinco hombres cuyo rostro puede distinguirse. De izquierda a derecha son: ceguera resultante de la enfermedad conocida como pénfigo, atrofia del globo ocular, leucoma corneal, amaurosis y enucleación.

En 1565 finalizó la construcción de un canal entre Antwerp y Bruselas. Los prohombres de esta última ciudad eligieron a Bruegel para que representara, en una serie de cuadros, la construcción del mencionado canal. Bruegel nunca terminó estos cuadros porque murió en 1569 cuando tenía cuarenta y cuatro años como máximo (la fecha de su nacimiento no es segura, pero suele situarse entre 1525 y 1530). El hecho de que se eligiera a Bruegel para este encargo sugiere que los vecinos de Bruselas, orgullosos de su canal y de la tecnología representada en el laborioso puerto de Antwerp, valoraban sus fieles retratos de la gente y de los artefactos de su mundo. Esta habilidad lo convierte en un guía fructífero, no sólo para el estudio de las tecnologías importantes de su época, sino también para el de ideas científicas fundamentales (tanto acertadas como erróneas) de aquellos tiempos lejanos.

El mantenimiento de la variabilidad genética de las poblaciones

El estudio molecular de los polimorfismos proteicos ha revelado la presencia de una gran variabilidad genética en las poblaciones. El significado evolutivo de esta variabilidad se interpreta de forma distinta por neutralistas y por seleccionistas

Antonio Fontdevila

El proceso de diferenciación que ha conducido a la diversidad biológica actual es una de las principales áreas de estudio de la biología. A Darwin se debe la formulación inicial del mecanismo de selección natural para explicar dicho proceso. El fundamento del mismo está en que la distinta constitución genética (genotipo) de los seres vivos determina una diferente capacidad de dejar descendientes (eficacia biológica) para cada organismo de una especie. Esto produce con el tiempo la selección en las poblaciones de aquellos genotipos con una mayor eficacia biológica, los cuales son los mejor adaptados a su ambiente. Puesto que el ambiente es cambiante, genotipos que estaban adaptados pueden dejar de estarlo y la selección favorece a otros nuevos. Este proceso conduce a cambios en los caracteres orgánicos que llegan a producir no sólo variaciones dentro de una población, sino que, al acumularse, pueden llevar a la formación, primero de razas y variedades, y, finalmente, de especies distintas.

La inmensa mayoría de los científicos actuales aceptan la teoría de la evolución biológica. Esto se debe no sólo a que ésta da una explicación coherente de la diversidad biológica, sino también a que permite la comprobación experimental de los cambios evolutivos. Los mecanismos evolutivos actuales no son distintos de los que actuaron en otras épocas y, por consiguiente, el estudio de los mismos en el laboratorio o en la naturaleza es válido para el conocimiento histórico del proceso evolutivo. La aparición brusca de nuevos genes, por mutación, es un fenómeno que se da tanto en la naturaleza como en nuestros laboratorios. Del mismo modo, la selección natural puede observarse también en trabajos experimentales y, de hecho, la selección ha sido practicada con éxito por los mejoradores de plantas y los criadores de ganado durante milenios.

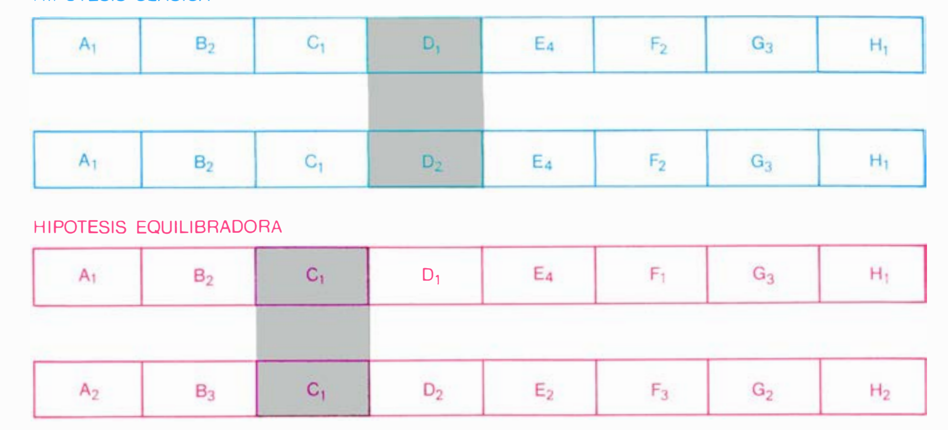
Desde el principio de la formulación de la teoría evolutiva quedó claro que las mutaciones eran la fuente principal de la nueva variabilidad genética. Sin embargo, pronto resultó evidente para el mismo Darwin que no sólo era necesario explicar la aparición de nueva variabilidad, sino que también era preciso que ésta se conservara de una generación a otra para que el proceso evolutivo continuara. Darwin no formuló una teoría genética coherente que explicara el mantenimiento de la variabilidad a través de la herencia. Los trabajos de Mendel, aunque contemporáneos a los suyos, le fueron desconocidos, así como a la mayoría de los naturalistas de la época. Cuando, a principios de este siglo, se redescubren los experimentos de Mendel, muchos investigadores comprenden que el mecanismo mendeliano de la herencia, en que los genes son unidades discontinuas que ocupan un lugar definido (locus, plural loci) en el cromosoma, permite abordar la explicación de dicho mantenimiento. Para ello, inician la importante tarea de estudiar genéticamente la variabilidad de las poblaciones naturales. Estos estudios han conducido a la formulación de dos teorías en cierto modo contrapuestas.

La primera (hipótesis clásica) considera que la selección natural tiene un papel básicamente conservador y purificador. Su autor, H. J. Muller, había observado en sus trabajos sobre mutaciones de *Drosophila* (la mosca del vinagre) que éstas eran generalmente perjudiciales (deletéreas) y que, por tanto, la selección natural las eliminaba. De este modo se conservaba la pureza de los genotipos básicos (tipos salvajes) de una población. La consecuencia de esta hipótesis es que en los organismos superiores, con una constitución cromosómica formada por pares de cromosomas homólogos (diploides), dichos genotipos presentan dos genes iguales (uno en cada cromosoma homólogo) en la mayoría de loci (homocigoto) y sólo ocasionalmente un locus tiene dos genes distintos (heterocigoto). Algunas veces una mutación es beneficiosa y entonces es seleccionada, sustituyendo al gen salvaje y pasando a ser ella el nuevo tipo salvaje homocigótico. La evolución procede, por tanto, mediante sustituciones de genes mutantes favorables.

Por el contrario, los trabajos de Th. Dobzhansky, también con *Drosophila,* le llevaron a la conclusión de que los genotipos de una población son básicamente heterocigóticos para la mayoría de sus loci. Esta segunda hipótesis en su forma más extrema explica el mantenimiento del polimorfismo genético porque si la selección natural favorece al heterocigoto, mantiene los genes de cada locus (alelos) en equilibrio estable (hipótesis equilibradora). La evolución se produce al modificarse el equilibrio génico debido a cambios ambientales. El mecanismo equilibrador permite mantener abundantes cantidades de variabilidad genética, la cual se mide por el número medio de loci heterocigotos (heterocigosis). La heterocigosis no sólo tiene valor adaptativo inmediato, sino que constituye una reserva de variabilidad que puede utilizarse cuando cambios ambientales futuros exijan nuevas adaptaciones. La hipótesis equilibradora considera que la distinción entre alelos salvajes y alelos mutantes es arbitraria. Todos los genes tienen un papel básico en la adaptación de las poblaciones. La hipótesis equilibradora tiene defensores entre los muchos discípulos de Dobz-

hansky en Estados Unidos y también entre muchos evolucionistas de la escuela británica, como E. B. Ford. Su visión de los seres vivos es parecida a la que Darwin tenía: consideran la gran diversidad de formas biológicas como una realidad producida por la gran variabilidad genética subyacente.

La mayor parte de la variabilidad genética se ha estudiado a través del análisis de los fenotipos. El fenotipo, o conjunto de caracteres observables de un organismo, es el resultado final de la acción génica interactuando con un ambiente dado y, en muchos casos, se trata de un pálido reflejo del genotipo. Muchos caracteres de gran valor evolutivo, como los morfológicos (tamaño, peso, forma, etcétera) o los etológicos (comportamiento sexual, competencia por el alimento, etcétera) tienen una gran influencia ambiental y/o son el resultado de complejas interacciones génicas. De ahí que sea difícil estudiar su variabilidad genética. Otros caracteres que determinan directamente la eficacia biológica, tales como la viabilidad o la fecundidad, resultan también de complejos procesos génicos. En estos caracteres es difícil detectar pequeñas diferencias genotípicas debido a la influencia desviante del ambiente, y son precisamente esas pequeñas diferencias en estos caracteres las que tienen mayor valor evolutivo en la teoría darwiniana. Por el contrario, los genes que producen grandes efectos susceptibles de fácil medida constituyen una minoría y no representan fielmente el genotipo total.



LAS HIPOTESIS CLASICA Y EQUILIBRADORA explican la variabilidad genética de un modo radicalmente diferente en los organismos diploides. La primera supone que la mayoría de loci *(A,B,C...)* tienen sus dos genes (alelos) iguales (homocigotos) y sólo en unos pocos *(D)* son distintos (heterocigotos). Según la hipótesis equilibradora la mayoría de los loci son heterocigotos.

El estudio clásico de la variabilidad genética depende de la presencia de la propia variabilidad. Cuando no se conoce el producto directo de la acción génica, sólo es posible detectar la presencia del gen a través de la variabilidad fenotípica que produce. Si todos los guisantes hubieran tenido el mismo color o la misma textura de su piel, ¿cómo hubiera podido Mendel detectar la presencia de los genes? Y, sin embargo, los genes responsables del color o de la textura hubieran estado allí, aunque en forma de homocigotos de un solo tipo. Los genes determinan todas las estructuras y todas las funciones de un organismo, pero no es fácil saber el número de genes que intervienen en una estructura o en una función dadas. La genética clásica está basada en el análisis de los loci variables (con dos o más alelos), pero apenas permite el estudio de aquellos loci sin variabilidad. De ahí la gran importancia que han tenido los estudios con mutantes en el establecimiento de las bases de la genética por parte de Morgan y sus discípulos en los años 20.

El soporte físico del material genético son los ácidos nucleicos: ácido desoxirribonucleico (ADN) en la mayoría de los seres vivos y ácido ribonucleico (ARN) en algunos virus. Un gen es una parte de ácido nucleico que determina, a través de un proceso complejo, el orden de los aminoácidos en la síntesis de las cadenas polipeptídicas, las cuales constituyen las proteínas. Los genes pueden dirigir también la síntesis de ARN que interviene directamente en el metabolismo celular. La función de las proteínas es básica, no sólo porque constituye un importante material estructural de los seres vivos, sino, sobre todo, porque ciertas moléculas proteicas, los enzimas, permiten que se lleven a cabo las reacciones metabólicas.

Para un genético, el estudio de las proteínas representa la posibilidad de estudiar funciones sencillas determinadas directamente por uno o pocos genes. De este modo, la relación entre fenotipo (proteína) y genotipo (gen proteico) se hace más clara. Para un genético de poblaciones, el estudio de la variabilidad proteica representa, además, la posibilidad de cuantificar la variabilidad genética sin estar condicionado "a priori" por la variabilidad o no del locus a estudiar. La estructura genética de una población es un conocimiento básico para entender el mecanismo evolutivo; por esta razón la posibilidad de separar los distintos variantes de un mismo enzima (isoenzimas) mediante técnicas físicas abrió a la genética evolutiva nuevas posibilidades de estudio.

Aplicando una corriente eléctrica a una mezcla de proteínas situadas en un soporte poroso (gel de acrilamida, almidón o agar, por ejemplo), éstas emigran hacia el polo de signo distinto del de su carga eléctrica superficial (electroforesis). Cada forma proteica tiene una carga característica que determina su velocidad de emigración. Si al cabo de un tiempo de emigración se corta la corriente, cada forma proteica se encuentra en una posición característica distinta en el soporte. La visualización de las proteínas se logra con una técnica de revelado químico específico para cada tipo proteico, apareciendo unas bandas características en el soporte. Esta técnica permite distinguir no sólo los distintos isoenzimas producto de loci diferentes, sino también reconocer los isoenzimas determinados por alelos de un mismo locus (aloenzimas o alozimas). Mediante el estudio estadístico de las frecuencias con que aparecen las bandas alozímicas en una población se puede hacer un recuento de la heterocigosis para cada locus y también estimar la heterocigosis media para un conjunto representativo de loci alozímicos. Este método de detección no depende de la presencia de variabilidad, ya que los loci polimórficos (es decir, aquellos loci con al menos dos alelos, el más abundante con una frecuencia inferior a 0,99) se detectan con la misma probabilidad que los monomórficos (esto es, aquellos loci con un alelo cuya frecuencia es superior a 0,99).

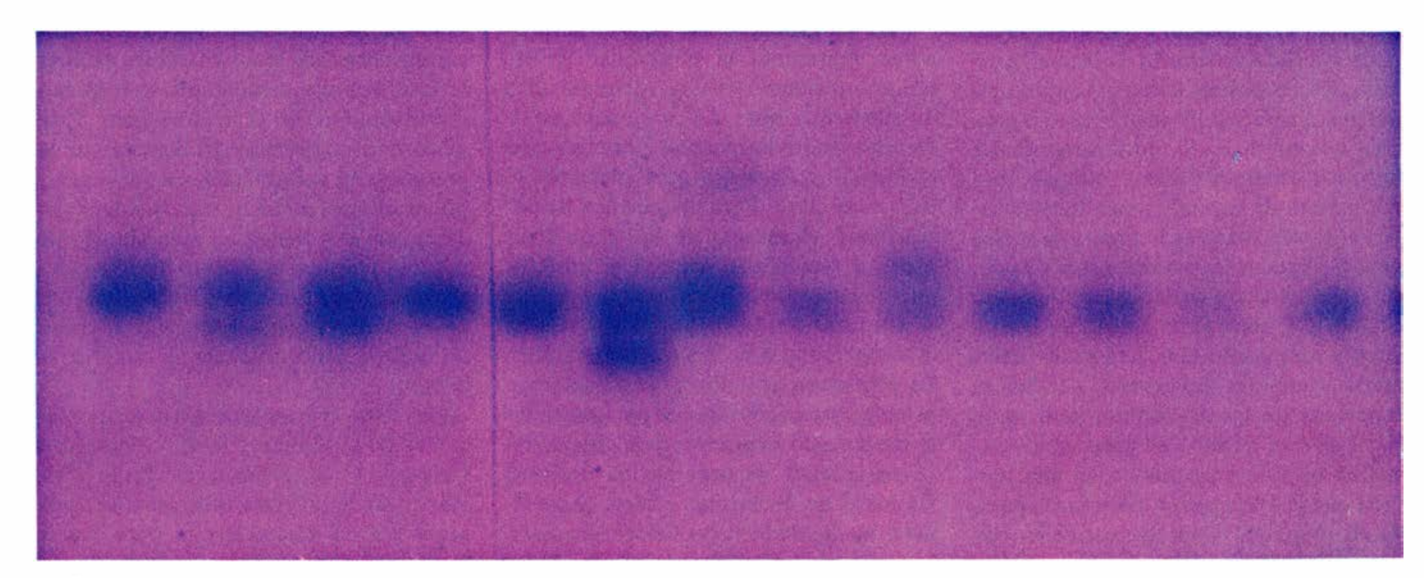
En 1966, varios genéticos de poblaciones (J. L. Hubby y R. C. Lewontin en Chicago, F. M. Johnson en Texas y H. Harris en Inglaterra) utilizaron por primera vez la electroforesis para medir la variabilidad genética. A partir de entonces, se observó que una gran proporción de loci alozímicos eran polimórficos (aproximadamente el 30 por ciento). Hoy en día se han ensayado más de 200 espe-

cies animales distintas y un gran número de especies vegetales, que abarcan grupos taxonómicos muy diversos en ambos reinos. El grado de polimorfismo alozímico en la mayoría de las poblaciones muestradas es muy elevado, lo mismo que su heterocigosis media (5 por ciento en vertebrados y 11 por ciento en invertebrados). Esta gran variabilidad observada parece favorecer el modelo equilibrador y los mecanismos evolutivos que lo mantienen. Sin embargo, existen algunas complicaciones teóricas en dicho modelo.

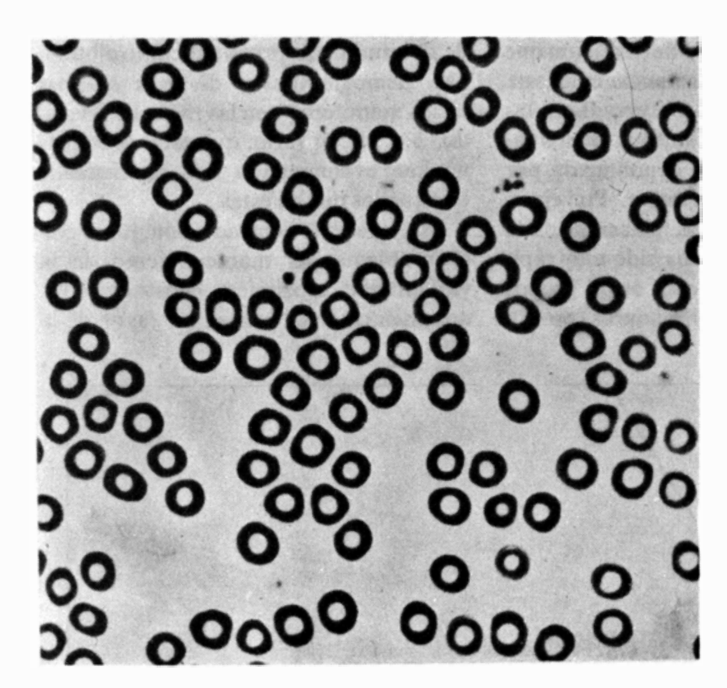
El modelo clásico admite sólo la presencia de unos pocos loci polimórficos. Estos mantienen algunos genes mutantes que confieren (especialmente en homocigosis), a los genotipos que los poseen, una baja eficacia biológica. Algunos de estos genes son letales (por ejemplo, la enfermedad Tay-Sachs en el hombre), otros sólo disminuyen la eficacia biológica (como la hemofilia); aunque todos ellos contribuyen a disminuir la capacidad reproductora de las poblaciones. Dobzhansky estimó que del 25 al 30 por ciento de todos los cromosomas de *Drosophila* contienen un gen letal o semiletal (con una viabilidad del 10 al 50 por ciento de lo normal). Muller y sus discípulos estimaron que en cada ser humano existe una media de 3 a 5 letales equivalentes (gen o conjunto de genes que producen la muerte en homocigosis). Muller llamó "lastre genético" a este conjunto de genes. Este lastre se produce en virtud de la aparición continua por mutación de genes deletéreos en las poblaciones; por eso suele denominarse lastre mutacional. Es evidente que si no fuera porque la mutación se produce con una frecuencia regular y constante, por gameto y generación, esto es, recurrente, la selección natural acabaría por eliminar los mutantes deletéreos de las poblaciones. El lastre mutacional se mantiene debido al equilibrio entre los mutantes que se van produciendo y los que la selección va eliminando. Sin embargo, el lastre mutacional no es el único estudiado por los genéticos. En realidad los defensores de la teoría del lastre mantienen que cualquier diferencia de eficacia biológica entre genotipos genera una disminución reproductora en las poblaciones, es decir, un lastre. Según esto, el mantenimiento de los polimorfismos equilibrados genera también un lastre denominado segregacional, ya que se debe a la inferior eficacia biológica de los homocigotos que se producen (segregan) en cada generación.

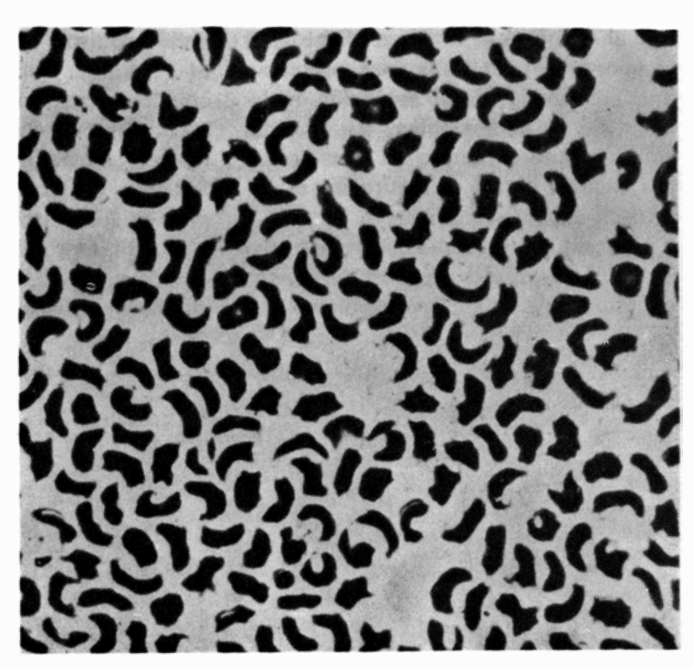
El mantenimiento de la gran variabilidad alozímica no puede explicarse por un mecanismo de equilibrio entre la selección en contra de alelos mutantes y la mutación recurrente. Este equilibrio no produce polimorfismos altos. Teóricamente puede explicarse mediante mecanismos equilibradores tales como la superioridad de los heterocigotos. Sin embargo, desde el primer momento se objetó que, dado el número elevado de loci polimórficos, el lastre segregacional debía ser excesivo para la supervivencia de las poblaciones. Supongamos que en un locus los homocigotos producen una disminución de capacidad reproductora del 1 por ciento. Si tomamos como cifra conservadora la de 10.000 genes proteicos en *Drosophila* y que un 30 por ciento de los loci (3000) son polimórficos en esta especie, entonces en el caso de independencia entre loci, la capacidad reproductora de una población media de *Drosophila* sería $(1-0,01)^{3000}=0,99^{3000}$. Esto supone que, para mantener la población estable, un individuo medio de esta población debe producir más de 800 billones de descendientes, cifra absolutamente absurda.

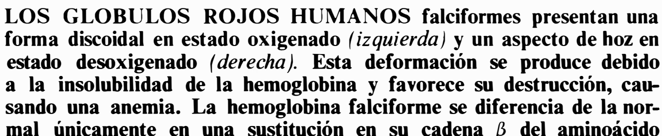
Existen muchas objeciones a la teoría del lastre segregacional. Quizá la objeción de más peso sea la que critica el aspecto relativo del cálculo del lastre, el cual no debería hacerse respecto al genotipo de mayor eficacia biológica (el heterocigoto). Imaginemos, por ejemplo, una población de homocigotos (monomórfica) para un locus alozímico. Supongamos que, por mutación, aparece en un momento dado un nuevo alozima que confiere al heterocigoto una mayor adaptación, por ejemplo, puede metabolizar mejor un sustrato nutritivo. ¿Cuál va a ser entonces el destino de esta población? ¿Va a disminuir su potencial reproductor? Lo más probable es que esta población alcance una mayor adaptación. En el nuevo equilibrio la población se habrá transformado en polimórfica y probablemente tendrá una mayor capacidad reproductora porque es capaz de utilizar mejor los sustratos nutritivos. Sin embargo, hemos pasado de una población monomórfica y, por consiguiente, sin lastre, a otra población polimórfica mantenida por superioridad del heterocigoto y con lastre segregacional. Esta paradoja resulta de considerar que los cálculos del lastre deben hacerse tomando como referencia una población cuyos



VARIABILIDAD PROTEICA en el locus enzimático leucín aminopeptidasa (Lap) de *Drosophila subobscura*, revelada por electroforesis en gel de almidón. En este locus los genotipos homocigóticos producen un solo tipo de alozima y originan por tanto una sola banda. Los heterocigotos originan dos tipos de alozimas y muestran dos bandas de distinta movilidad en el zimograma. Las posiciones 1, 4, 5, 8, 10, 11, 12 y 13, contando de izquierda a derecha, corresponden a homocigotos para la alozima Lap^{100}. Las posiciones 2 y 6 son heterocigotos Lap^{100}/Lap^{69}; la posición 3 es un heterocigoto Lap^{100}/Lap^{86}; la posición 7 es un heterocigoto Lap^{100}/Lap^{111} y la posición 9 es un heterocigoto para Lap^{100}/Lap^{118}.







LOS GLOBULOS ROJOS HUMANOS falciformes presentan una forma discoidal en estado oxigenado *(izquierda)* y un aspecto de hoz en estado desoxigenado *(derecha).* Esta deformación se produce debido a la insolubilidad de la hemoglobina y favorece su destrucción, causando una anemia. La hemoglobina falciforme se diferencia de la normal únicamente en una sustitución en su cadena β del aminoácido ácido glutámico por valina. El responsable de esta cadena β anormal es un gen mutante que en estado homocigótico produce una grave anemia, que generalmente acarrea la muerte prematura. Los heterocigotos presentan una forma leve de anemia y, en las zonas palúdicas, son menos susceptibles al ataque del paludismo que los homocigotos normales. Este caso de superioridad del heterocigoto produce un equilibrio génico.

individuos son todos del genotipo de mayor eficacia biológica (en el caso del lastre segregacional, el heterocigoto). Si en el ejemplo anterior se hubiera tomado como referencia la población monomórfica original, la aparición de un mutante heterocigótico hubiera producido un aumento de capacidad reproductora, o sea, un lastre negativo.

B. Wallace ha indicado que en los caracteres polimórficos la selección no se ejerce de un modo drástico y constante (selección dura). Los recursos naturales (alimento, lugares de cría, presencia de una pareja, etcétera) son frecuentemente los que determinan la capacidad reproductora de una población y no los genotipos que se hallan presentes. La selección elige los genotipos que presentan mayor eficacia biológica de entre todos los que se producen en cada generación; ahora bien, la capacidad reproductora de las poblaciones suele ser siempre máxima, y, por consiguiente, no genera prácticamente lastre (selección blanda).

Uno de los aspectos más interesantes del concepto de selección blanda es que la eficacia biológica de un genotipo no tiene por qué ser un parámetro constante. Diversos trabajos experimentales realizados con alozimas y caracteres morfológicos han revelado que la respuesta de un genotipo a la selección es inversamente proporcional a su frecuencia y afecta a componentes de la eficacia biológica tan dispares como viabilidad, o capacidad de llegar a la madurez, y eficacia en el apareamiento. Otros trabajos han demostrado que el valor adaptativo de un genotipo varía cuando éste ocupa distintas condiciones ecológicas.

El mantenimiento de la variabilidad genética mediante mecanismos de selección blanda elimina el problema del lastre planteado por la selección dura. Sin embargo, Kimura y sus colaboradores han desarrollado una hipótesis teórica que explica la presencia de variabilidad alozímica exclusivamente debida a la mutación y a los cambios aleatorios de las frecuencias génicas causados por el tamaño finito de las poblaciones (deriva genética). Esta teoría se ha denominado neutralista porque considera neutro el polimorfismo alozímico con respecto a la selección y, por tanto, elimina también el problema del lastre.

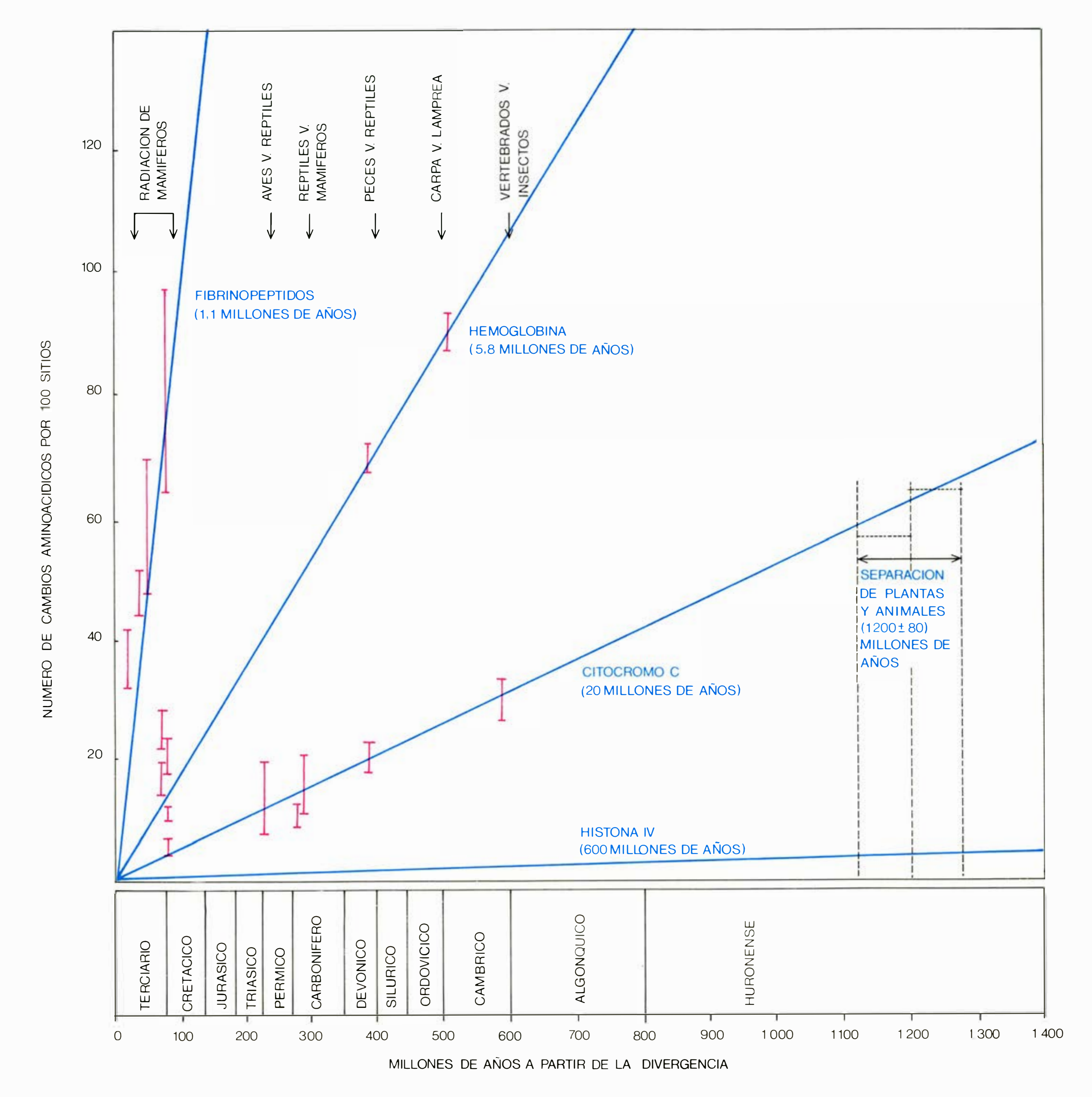
La genética molecular ha descifrado que las mutaciones génicas determinan cambios estructurales en el ADN, que se traducen en cambios paralelos en las proteínas. Muchos de estos cambios proteicos consisten en sustituciones de un aminoácido por otro. Por ejemplo, el mutante que produce anemia falciforme en el hombre provoca en la hemoglobina la sustitución del aminoácido ácido glutámico por valina. Las mutaciones que provocan sustituciones de aminoácidos se producen continuamente en la evolución de los seres vivos y muchas consiguen fijarse en homocigosis. Algunas de estas sustituciones producen los cambios en la carga superficial de las proteínas que se detectan mediante las técnicas de electroforesis. Si estudiamos la secuencia de aminoácidos de una misma proteína en distintos organismos encontramos distinto número de sustituciones. Es lógico que cuanto más alejadas evolutivamente estén dos especies, mayor número de aminoácidos diferentes contendrá la misma proteína en ambas. Utilizando estas diferencias moleculares se pueden construir árboles filogenéticos.

Uno de los argumentos más fuertes en favor de la teoría neutralista de la evolución, puesto de manifiesto por M. Kimura, del Instituto de Genética de Mishima (Japón), es la aparente constancia del número de sustituciones de aminoácidos por año en líneas filogenéticamente separadas. Así, por ejemplo, las cadenas alfa y beta de la hemoglobina están codificadas por dos genes que proceden de un gen que se duplicó hace unos 450 millones de años. Estas cadenas polipeptídicas han venido experimentando sustituciones de aminoácidos en distintas líneas de vertebrados, a una frecuencia sorprendentemente constante en todas ellas: aproximadamente, una sustitución por cien aminoácidos cada seis millones de años. Es difícil imaginar en la evolución de los vertebrados cómo líneas que han estado sometidas a condiciones adap-

tativas tan distintas como las que conducen al hombre, a la carpa, al ratón, al conejo, al caballo o a la vaca, presentan una constancia en la evolución molecular de la hemoglobina. Esta misma constancia se encuentra en otras proteínas, aunque la tasa de sustitución es distinta en cada una de ellas.

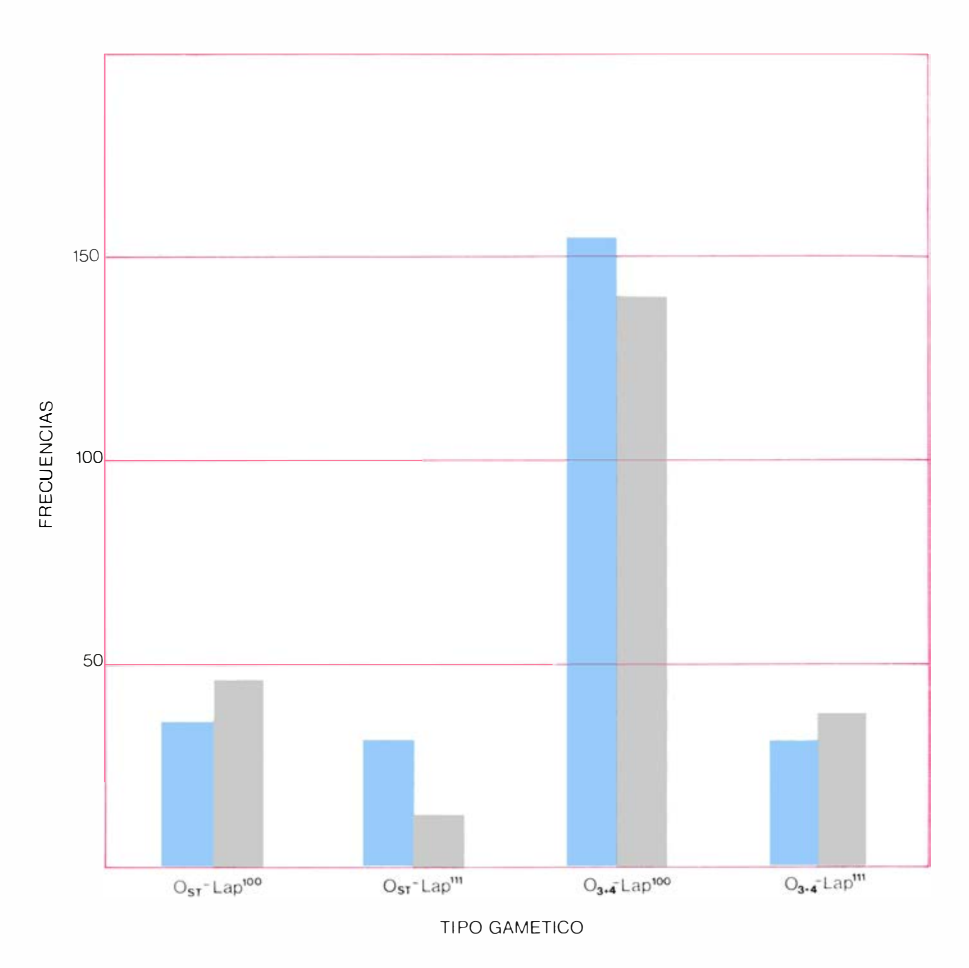
La tesis neutralista explica este hecho porque las sustituciones de aminoácidos no están sujetas a la selección natural (son neutras), y entonces la tasa de sustitución equivale a la tasa de mutación que es constante. Esta constancia contrasta con la discontinuidad observada en las tasas o frecuencias evolutivas de los caracteres morfológicos, denominada por Simpson evolución cuántica. Por ejemplo, en los vertebrados placentarios la evolución morfológica ha sido muy rápida en comparación con otros grupos no placentarios, como los anuros (ranas). El estudio comparativo de la evolución de determinadas proteínas como albúminas, hemoglobinas y diversos enzimas en los mamíferos y en las ranas ha revelado, por el contrario, que estas proteínas no han evolucionado más rápidamente en aquéllos que en éstas.

Esta aparente contradicción, junto con el problema del mantenimiento de la variabilidad alozímica, constituyen los argumentos más fuertes en favor de la



LA TASA DE EVOLUCION DE LAS PROTEINAS se deduce representando el número de sustituciones diferenciales de aminoácidos *(en ordenadas)* producidas desde el tiempo geológico *(en abscisas)* en que dos especies se han diferenciado a partir de un antepasado común; por ejemplo, el punto de divergencia entre la carpa y la lamprea se produjo a principios del Ordovícico, hace unos 500 millones de años. Los segmentos verticales indican el intervalo experimental de variación de dichas sustituciones. La tasa de evolución es proporcional a la pendiente de la recta y viene expresada por el período evolutivo unitario *(entre paréntesis debajo de cada proteína)*, que indica el número medio de años necesarios para sustituir un 1 por ciento de aminoácidos. El período evolutivo unitario es constante para cada proteína, pero distinto de unas proteínas a otras. (Según R. E. Dickerson, "The structure and history of an ancient protein", SCIENTIFIC AMERICAN, abril 1972.)

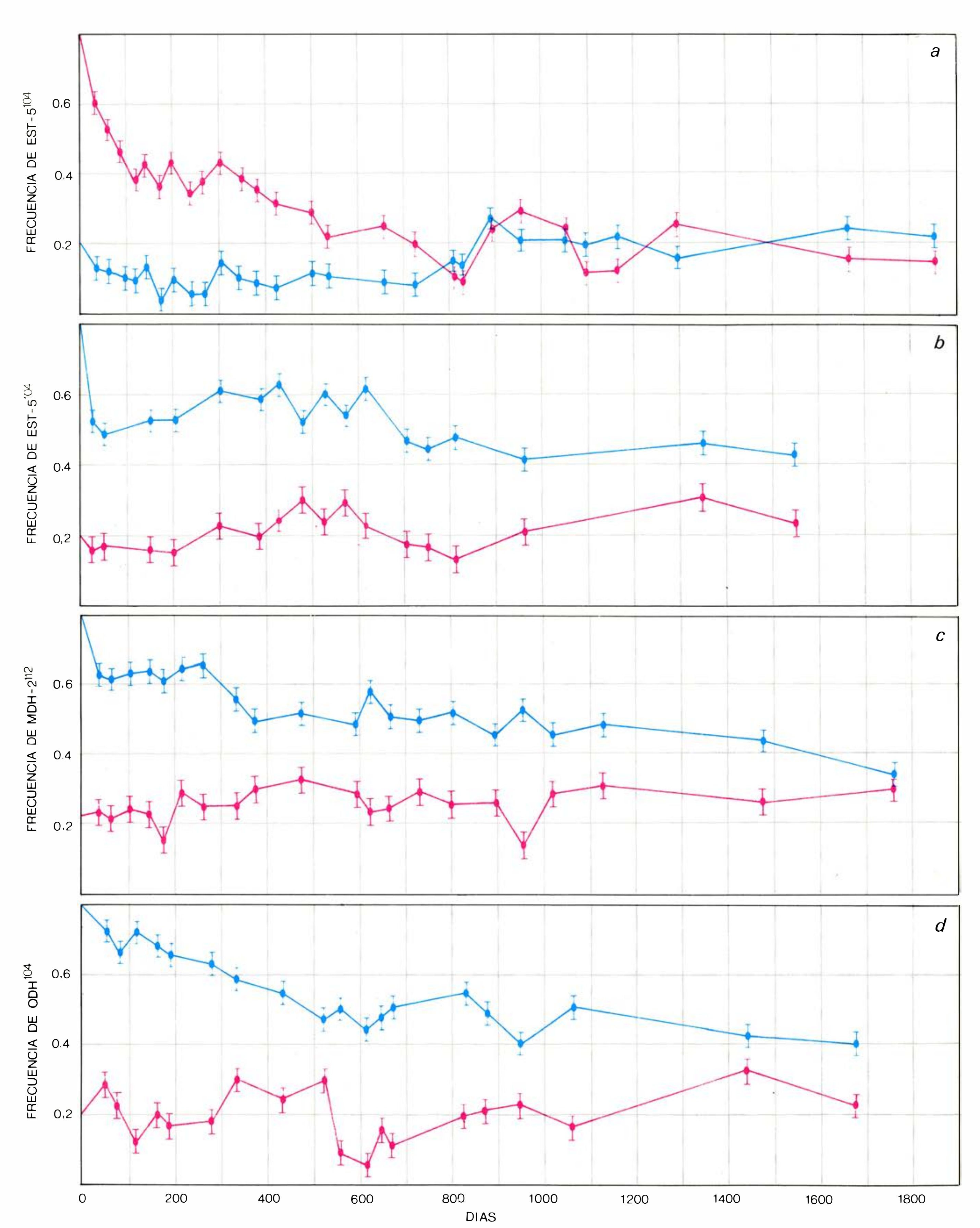
teoría neutralista. Por lo que respecta a la constancia de tasas evolutivas moleculares, los seleccionistas llaman la atención sobre dos puntos principales. Primero, las tasas de evolución son diferentes para distintas proteínas. Así, es mayor para los fibrinopéptidos que para la hemoglobina, y mayor en ésta que en el citocromo C. Estas diferencias están relacionadas con la función específica de dichas proteínas: cuanto más estricta es la función de una proteína, mayor resistencia opone ésta a las sustituciones de aminoácidos y, por tanto, menor es su tasa de evolución molecular. Por consiguiente, la evolución molecular de las proteínas está sujeta a la selección, la cual sólo permite aquellas sustituciones que mejoran o al menos no perjudican su función. Existen lugares en la proteína que no pueden modificarse mediante sustituciones porque éstas producen alteraciones tan drásticas que la proteína deja de ser funcional. En otros lugares sólo son permisibles determinadas sustituciones. El segundo punto de los seleccionistas es que muchas proteínas podrían evolucionar adaptándose al medio interno de la célula. Esta adaptación se realiza de un modo conservador y gradual: la proteína retiene su funcionalidad con el cambio experimentado en dicho medio. Un ejemplo de adaptación al medio interno lo constituye el hecho de que la única sustitución diferencial entre el hombre y el gorila en la cadena beta de la hemoglobina parece ser específica de cada especie. Ninguna de las sustituciones encontradas en la cadena beta de la hemoglobina humana se encuentra en la región de dicha sustitución diferencial. La más próxima produce serias alteraciones patológicas porque afecta a la estructura del grupo activo hemo y a la zona de contacto entre las cadenas alfa y beta. Los cambios genéticos no producen en muchos casos cambios fenotípicos (morfológicos) inmediatos, ya que el fenotipo no es la expresión directa del genotipo. Sólo cuando se presenta un determinado conjunto de cambios ambientales y genéticos que interactúan entre sí, aparece bruscamente un nuevo fenotipo y se produce un caso de evolución cuántica. En el hombre existe un tipo de hemoglobina (Zurich) que presenta en el lugar 63 de su cadena beta una sustitución del aminoácido histidina normal por arginina. Esta hemoglobina no es anormal, a menos que la persona ingiera sulfanilamida. Entonces los glóbulos rojos lisan y se produce una anemia hemolítica. Este caso constituye un ejemplo de la aparición brusca de un nuevo fenotipo. La presencia de evolución morfológica discontinua podría ser el resultado de procesos análogos.



DESEQUILIBRIO DE LIGAMIENTO entre inversiones del cromosoma O y el alozima del locus genético leucín aminopeptidasa (Lap) en la población natural El Pedroso (Santiago de Compostela). Las diferencias de frecuencias entre los valores observados *(azul)* y los valores esperados *(gris)* son estadísticamente significativas. Los tipos gaméticos O_{3+4}-Lap^{100} y O_{st}-Lap^{111} se presentan en frecuencias más altas de lo que cabría esperar si la asociación fuera debida al azar.

Sin embargo, muchos seleccionistas piensan que gran parte de las sustituciones proteicas se mantienen como una respuesta adaptativa directa al medio externo (ambiente). Esto contrasta con la hipótesis neutralista, que no atribuye a la selección natural, sino a la mutación y a los cambios aleatorios en las frecuencias génicas, el papel básico en el mantenimiento de los polimorfismos proteicos. Según esto, la teoría neutralista predice una falta de correlación génica entre distintas poblaciones y también entre poblaciones y factores del ambiente (clinas). Los cambios aleatorios de las frecuencias génicas debidos básicamente al tamaño reducido del número de genes que se transmiten de una a otra generación (deriva genética), han de producir polimorfismos y fijaciones génicas de muy distinto signo en diferentes poblaciones. Sin embargo, muchas distribuciones génicas presentan correlaciones altas entre poblaciones, muy separadas geográficamente, de diversas especies de *Drosophila (D. pseudoobscura, D. willistoni, D. obscura, D. subobscura)* y de otros organismos, como el ratón doméstico, por ejemplo. Ayala y otros autores han puesto de manifiesto que generalmente los loci monomórficos lo son para el mismo alelo en todas las poblaciones, y que los loci polimórficos tienen el alelo más frecuente común a todas las poblaciones. Esto descarta, aparentemente, una explicación neutralista de dichas distribuciones, a menos que supongamos una pequeña migración entre poblaciones. La migración es un parámetro básico para interpretar las distribuciones génicas en el espacio; sin embargo, no es fácil de medir y, en muchos casos, desconocemos su alcance real.

La distribución geográfica de asociaciones entre alelos de distintos loci está siendo muy estudiada actualmente porque parece no presentar tantos problemas de discriminación entre ambas hipótesis. En primer lugar, los genes no funcionan solos y es lógico que se seleccionen determinados grupos de genes (tipos gaméticos), que se trànsmiten juntos debido al alto valor selectivo del conjunto. Esto conduce a un desequilibrio de ligamiento o de fase ga-



EL EFECTO DE LA SELECCION en poblaciones experimentales se mide mediante los cambios temporales de las frecuencias génicas. En la figura se han representado dichos cambios en poblaciones de *Drosophila pseudoobscura* para los loci alozímicos esterasa-5 (Est-5) *(a, b)*, málico deshidrogenasa-2 (Mdh-2) *(c)* y octanol deshidrogenasa (Odh), *(d)*. Se observa *(a)* que las frecuencias génicas para el alelo Est-5^{104} de las poblaciones *A (línea roja)* y *B (línea roja)* parten de valores iniciales muy distintos (0,80 y 0,20); sin embargo, al cabo de unas 30 generaciones (aproximadamente 900 días) dichas frecuencias convergen en la frecuencia de equilibrio de 0,20. Las poblaciones A_1 *(línea azul)* y B_1 *(línea azul)*, fundadas a partir de las mismas cepas iniciales que las *A* y *B*, presentan cambios análogos en las frecuencias del alelo Est-5^{104} *(b)*. No obstante, en este caso la tendencia de las frecuencias es hacia un nuevo equilibrio algo distinto (0,30) del anterior. Se registran cambios *(c)* en las frecuencias del alelo Mdh-2^{112} hacia valores de equilibrio de 0,30. Y cambios en las frecuencias alélicas de Odh^{104} hacia valores de equilibrio análogos a los anteriores (0,30) *(d)*. Los segmentos verticales indican los valores de error típico de las frecuencias.

mética en el que la distribución no es al azar, predominando unos tipos gaméticos respecto a otros. En segundo lugar, Lewontin y otros autores han puesto de manifiesto que la presencia de desequilibrio de ligamiento del mismo signo en diversas poblaciones no puede explicarse por deriva genética y debe atribuirse a la selección natural. En este caso, la migración debería ser muy fuerte para uniformizar el desequilibrio generado por deriva en todas las poblaciones. Ahora bien, con grandes tasas de migración todas las poblaciones funcionarían como una sola y su tamaño sería entonces demasiado grande para que la deriva genética pudiera generar desequilibrio de ligamiento. En tercer lugar, la formación de bloques de genes supone la no independencia entre loci, lo cual disminuye mucho el lastre genético y, por consiguiente, elimina esta objeción para el mantenimiento de los polimorfismos alozímicos.

Todavía no son muy abundantes los datos sobre desequilibrio de ligamiento. Zouros y Krimbas han encontrado en *Drosophila subobscura* pruebas de desequilibrio de ligamiento entre los loci alozímicos xantin deshidrogenasa (Xdh) y aldehído oxidasa (Ao). Los autores atribuyen este desequilibrio a la selección, puesto que es del mismo signo y magnitud en dos poblaciones aisladas. También se han encontrado casos de asociación del mismo signo entre inversiones y alozimas. Una inversión es un segmento cromosómico cuya posición dentro del cromosoma sufre un giro de 180 grados con respecto a otros segmentos análogos. Las inversiones evitan la recombinación entre los loci incluidos en ellas, de modo que éstos funcionan como un bloque de genes. El hecho de que se mantengan inversiones en la mayoría de las poblaciones naturales de distintas especies de *Drosophila* se ha atribuido al efecto de la selección natural que favorece a dichos bloques coadaptados de genes. Krimbas y Loukas en Grecia, Charlesworth en Inglaterra, Prevosti en diversos lugares de Europa y el autor en el NO de España han encontrado desequilibrio de ligamiento del mismo signo entre alozimas del locus leucin aminopeptidasa (Lap) e inversiones del cromosoma O en *Drosophila subobscura*. Esta constancia en el signo del desequilibrio entre poblaciones de muy diverso origen geográfico parece favorecer a la selección natural como principal causante del mismo.

Otros trabajos con poblaciones naturales, principalmente con *Drosophila*, aunque también con otros organismos, como peces, caracoles y gramíneas, demuestran la presencia de clinas en distintos polimorfismos proteicos. Algunos autores han encontrado también correlaciones entre variabilidad proteica y variación estacional.

Sin embargo, muchos de los trabajos con poblaciones naturales tropiezan con dos problemas principales. El primero radica en la dificultad de medir determinados parámetros poblacionales tales como tasa de migración, tamaño de población, estructura demográfica, etcétera. Estos parámetros ecológicos determinan las predicciones de otros parámetros genéticos como la heterocigosis o el número de alelos de un polimorfismo. El desconocimiento de dichos parámetros ecológicos hace muy difícil en muchos casos decidir si una determinada distribución observada puede o no atribuirse a la selección natural.

El segundo problema consiste en demostrar de un modo directo que la selección natural es la responsable de un determinado polimorfismo proteico. Aun cuando el muestreo de las poblaciones naturales sugiera que la selección natural es el factor explicativo de un polimorfismo, ¿cómo podemos estar seguros de que está actuando en el locus estudiado? ¿no podría actuar en otros loci estrechamente ligados? Para resolver esta cuestión es preciso conocer las relaciones ecológicas entre el organismo y su ambiente, es decir, su nicho ecológico.

Con objeto de controlar algunos parámetros en las poblaciones, muchos investigadores han acudido al estudio experimental de poblaciones mantenidas en el laboratorio. Estas poblaciones artificiales se fundan generalmente con un número dado de individuos que se introducen en cajas aisladas. En estas cajas de poblaciones, los organismos se reproducen y, si las condiciones ecológicas (alimentación, humedad, temperatura, etcétera) son las adecuadas, la población se mantiene durante un número indefinido de generaciones. Las condiciones ambientales pueden controlarse y manipularse, de tal modo que los cambios génicos observados son más fácilmente atribuidos o no a la selección que opera en estas poblaciones.

El autor, y sus colegas Francisco J. Ayala y John McDonald de la Universidad de California en Davis (Estados Unidos), y Josefina Méndez, de la Universidad de Santiago de Compostela (España) han realizado una serie de experimentos con poblaciones de laboratorio y distintos sistemas alozímicos de *Drosophila pseudoobscura:* Esterasa - 5 (Est-5), Málico deshidrogenasa-2 (Mdh-2) y Octanol deshidrogenasa (Odh). Estos experimentos han revelado cambios génicos significativos en todos los sistemas estudiados. El locus Est-5 parece haber alcanzado el equilibrio génico al cabo de unas 30 generaciones de la fundación de la población y se ha mantenido en el equilibrio por espacio de otras 30 generaciones. Los loci Odh y Mdh-2 presentan también una tendencia significativa hacia un equilibrio génico. La tendencia general de todas las cajas de poblaciones hacia equilibrios génicos, que, además, se aproximan a las frecuencias de equilibrio encontradas en la naturaleza, sugiere que la selección es responsable, al menos en parte, de estos cambios de frecuencias. Sin embargo, existe la posibilidad de que la selección actúe fundamentalmente en otros loci ligados y que los loci observados sean marcadores neutros o casi neutros.

Las moscas fundadoras de las cajas *(A, B)* se obtuvieron mediante cruzamientos consanguíneos a partir de descendencias (cepas) de hembras individuales fecundadas, capturadas en la naturaleza. Con estas mismas cepas, y después de aproximadamente 12 generaciones (un año) de que éstas fueran mantenidas en el laboratorio, se fundaron dos nuevas cajas (A_1B_1), réplicas de las anteriores. Los genotipos han respondido nuevamente a la selección evolucionando las frecuencias génicas en el mismo sentido que en las poblaciones originales *(A, B)*. Sin embargo, estas cajas presentan una tendencia significativa hacia un nuevo equilibrio distinto del anterior, en el sentido de que la selección parece favorecer ahora al alelo Est-5^{104} más que antes.

Estas diferencias selectivas para un mismo alozima sugieren que otros loci interactúan con el locus alozímico. El mantenimiento prolongado en el laboratorio de las cepas fundadoras de las cajas A_1B_1 en un ambiente de alta competitividad y densidad (cultivos masales) puede haber promovido la selección de genes modificadores del gen Est-5^{104}, que aumenten su capacidad de explotar el ambiente en dichas condiciones. Estos resultados apoyan la idea de que, si bien el locus Est-5 parece estar sujeto a la selección, su valor adaptativo puede ser modificado con relativa facilidad.

Para detectar diferencias en algunos componentes de la eficacia biológica entre distintos genotipos del locus Est-5, J. Méndez y el autor realizaron una serie de experimentos que midieron la viabilidad, la fecundidad y la capacidad de apareamiento. La viabilidad huevo-adulto no mostró diferencias significativas entre genotipos. Por el contrario, la fecundidad resultó ser más alta en las hembras homocigóticas para Est-5^{104}, mos-

trando las hembras heterocigóticas una fecundidad intermedia. Estos resultados no explican, sin embargo, el distinto comportamiento selectivo entre poblaciones *A*, *B* y A_1B_1 al no encontrarse diferencias entre ellas para estos componentes. Por el contrario, el experimento de capacidad de apareamiento reveló diferencias entre las poblaciones según su origen. Así, cuando los genotipos proceden de las cajas *A*, *B*, las frecuencias de apareamiento corresponden a las esperadas según las frecuencias iniciales, aunque existe una cierta ventaja de los machos de genotipo Est-5^{100}. Sin embargo, cuando los genotipos son originarios de las cajas A_1B_1 las frecuencias observadas de apareamiento se desvían de las esperadas para frecuencias extremas. Esta desviación significativa se debe a que la capacidad de apareamiento de un genotipo está inversamente correlacionado con su frecuencia.

Esta es la primera vez que se ha encontrado, en un sistema alozímico, una selección dependiente de las frecuencias para capacidad de apareamiento. La presencia de diferencias de eficacia biológica entre genotipos sugiere que la selección natural puede actuar en el sistema Est-5. Los resultados indican, además, que la selección natural actúa favoreciendo a distintos genotipos en diferentes componentes de la eficacia biológica. Así, el genotipo homocigótico para Est-5^{104} es favorecido en las hembras debido a su mayor fecundidad, mientras que en machos parece tener una capacidad de apareamiento menor. Estas acciones contrapuestas de la selección para un mismo genotipo explicarían el equilibrio encontrado en las cajas de poblaciones. Por otra parte, los resultados indican también que la capacidad de apareamiento se ha modificado, pasando a depender de las frecuencias. Esto explicaría el nuevo equilibrio de las poblaciones A_1 y B_1. En ninguno de los componentes estudiados de la eficacia biológica se ha observado una superioridad de los heterocigotos; por el contrario, éstos presentan valores intermedios. La conclusión a que se llega es que la selección se ejerce mediante mecanismos "blandos" (eficacia biológica variable en distintas fases del ciclo vital o dependiente de las frecuencias génicas), que no generan lastre y permiten mantener altos valores de polimorfismo. Con todo, estas pruebas son indirectas y no nos dicen con certeza si la selección actúa directamente sobre el sistema alozímico.

Gillespie y Kojima, de la Universidad de Texas, fueron los primeros en apuntar la posibilidad de que el grado de polimorfismo de un sistema proteico estuviera relacionado con su función metabólica. Atacaban el problema, pues, de una manera directa. Para ellos, y para otros, el polimorfismo es bajo en un conjunto de enzimas (grupo I) que actúan sobre sustratos fisiológicos definidos en ciertos procesos intracelulares, tales como la biosíntesis, los pasos del metabolismo intermedio o los ciclos metabólicos de producción de energía. Por el contrario, otro grupo (II) de enzimas que actúa sobre sustratos externos más inespecíficos presenta un mayor polimorfismo, que posiblemente refleja la variabilidad ambiental. Algunos autores subdividen el grupo I en dos subgrupos, según el papel regulador o no de los enzimas en los pasos limitantes del metabolismo. Los enzimas reguladores tienen por misión proteger al organismo de las fluctuaciones ambientales; de ahí que poseen mayor flexibilidad, o variabilidad, que los enzimas no reguladores, que actúan independientemente de los cambios ambientales. Aunque estas hipótesis pueden ayudar a la explicación del polimorfismo proteico, tropiezan con serios problemas. Uno de ellos consiste en la dificultad para asignar muchos enzimas a uno u otro grupo debido al desconocimiento de su función "in vivo".

Otro problema estriba en las diferencias en el polimorfismo de un mismo enzima según la especie que se considere. Este es el caso de *D. buzzatii* estudiada por Barker en Australia y por el autor y sus colaboradores en la Universidad de Santiago de Compostela. Esta especie pertenece al grupo *repleta*, subgrupo *mulleri*, y presenta como las demás especies del grupo una de las hererocigosis medias más bajas dentro del género *Drosophila*. Sin embargo, el locus alcohol deshidrogenasa (Adh) es altamente polimórfico en *D. buzzatii*, con una heterocigosis (H) de alrededor de 0,50 que contrasta con las bajas heterocigosis de las demás especies del grupo. Por el contrario, el locus leucin aminopeptidasa (Lap) es prácticamente monomórfico en dicha especie, mientras que es bastante polimórfico en las demás. Ambos enzimas han sido clasificados dentro del grupo II y estos resultados demuestran la relativa arbitrariedad de esta clasificación.

SELECCION SEXUAL DEPENDIENTE DE LAS FRECUENCIAS GENICAS para el locus Esteresa-5 de *Drosophila pseudoobscura*. En el gráfico se representa la regresión logarítmico-lineal del cociente de las frecuencias de apareamiento (razón de salida; *en ordenadas*) entre los dos genotipos homocigóticos para los alozimas Est-5^{100} y Est-5^{104}, sobre el cociente de frecuencias iniciales (razón de entrada; *en abscisas*) para los mismos genotipos. Si los genotipos se aparean al azar, la razón de entrada es igual a la de salida y los puntos se sitúan sobre la diagonal (caso de las cajas *a* y *b*, *triángulos*). Cuando hay selección sexual los puntos se desvían de la diagonal. Tal es el caso de los genotipos de las cajas A_1 y B_1 (*círculos*), en el que, además, el sentido de las desviaciones de las frecuencias de apareamiento está en relación inversa a las frecuencias iniciales.

Otra especie, *D. melanogaster,* presenta una heterocigosis alta para el locus Adh. El polimorfismo natural de Adh en esta especie se ha relacionado con factores ambientales, como la temperatura o la latitud, habiéndose encontrado clinas geográficas por diversos autores. Los estudios bioquómicos realizados con el locus Adh en esta especie demuestran que existen diferencias significativas entre los dos alelos principales del sistema (Adh^F y Adh^S). La alozima Adh^F es más activa pero menos estable que Adh^S y ambos alozimas difieren también en su especificidad de sustrato, que es un carácter peculiar de los enzimas, y en otras propiedades bioquímicas. Pero lo más interesante es que C. L. Vigue, F. M. Johnson y A. Powell, de la Universidad del Estado de Carolina del Norte en Raleigh (Estados Unidos), han demostrado que Adh^S presenta más actividad después de un tratamiento de calor (40 grados Celsius) que Adh^F y es precisamente aquella alozima (Adh^S) la que se encuentra en alta frecuencia (aproximadamente 0,90) en la zona sur de una clina que presenta *D. melanogaster* al este de los Estados Unidos. Por el contrario, Adh^S alcanza sólo una frecuencia de 0,50 al norte de la clina. Estos estudios con *D. melanogaster* sugieren que el sistema Adh es adaptativo y que descompone los alcoholes que se encuentran en el sustrato larvario de esta especie. Los resultados experimentales apoyan la idea de que el polimorfismo encontrado en Adh no sólo es adaptativo, sino que se mantiene mediante mecanismos equilibradores, como la selección dependiente de las frecuencias o los cambios de valor selectivo en nichos múltiples.

La demostración de que en uno o en unos pocos sistemas proteicos se presentan diferencias "in vitro" en las propiedades bioquímicas de sus alelos no constituye ninguna demostración general del valor selectivo del polimorfismo proteico. Por otra parte, no existe seguridad de que dichas diferencias sean detectadas en condiciones fisiológicas "in vivo". Por otro lado, estos casos podrían representar la excepción más que la regla. Aunque el caso de Adh es posiblemente el más documentado debido a la concordancia de los datos que se poseen, tanto en poblaciones experimentales y naturales como en condiciones "in vitro", no es el único. Otros estudios en peces demuestran también análogas correlaciones entre actividad diferencial "in vitro" y frecuencia alozímica en clinas.

La controversia entre neutralistas y seleccionistas se ha centrado básicamente en los polimorfismos proteicos detectables electroforéticamente. Al principio se pensó que esta fracción podía ser representativa de todo el genoma. Hoy en día se tienen serias dudas. Posiblemente, sólo entre un 20 y un 25 por ciento de todos los cambios de aminoácidos producen cambios en la carga neta de las proteínas. Pero, además, no estamos seguros de que los cambios evolutivos importantes se produzcan en la fracción de ADN donde se encuentran los genes que dirigen la síntesis de proteínas (genes estructurales). En concreto, ¿son los cambios en genes estructurales los determinantes de los cambios evolutivos iniciales que conducen a la formación de nuevas especies? En bacterias, los primeros cambios génicos adaptativos ante una nueva fuente energética se producen generalmente en genes que regulan la acción de los genes estructurales, por lo que reciben el nombre de genes reguladores; sólo posteriormente se producen cambios en genes estructurales. Disponemos ya de datos que sugieren que las reorganizaciones cromosómicas en organismos superiores pueden alterar la expresión de los genes modificando funciones antiguas en otras nuevas. Entre estas reorganizaciones tenemos, además de los inversiones, las translocaciones (intercambio de fragmentos entre cromosomas), las duplicaciones de segmentos cromosómicos, las deficiencias o pérdidas de fragmentos y las fusiones entre cromosomas. Estos cambios pueden ser rápidos y no estar acompañados por cambios paralelos en los genes estructurales alozímicos.

Además del ADN que contiene genes estructurales, existen muchas otras clases de ADN. Se ha sugerido que menos de un uno por ciento del ADN de un organismo superior está constituido por genes estructurales. El resto, es decir, la mayor parte del ADN tiene otras funciones importantes, entre las que se encuentra la regulación génica. Esta función, descubierta por Jacob y Monod en la bacteria *Escherichia coli,* está siendo estudiada en organismos superiores y existen ya casos descritos de genes reguladores en diversos de estos organismos, como maíz o *Drosophila.* Sin embargo, todavía conocemos poco de la estructura genética del ADN regulador y mucho menos de su variabilidad. Esta variabilidad puede ser decisiva a la hora de estudiar la evolución de los seres vivos y, a lo mejor, mucho más importante que la variabilidad proteica estructural para la comprensión de los cambios evolutivos. En este sentido, si la variabilidad proteica resultase ser neutra no por esto la evolución dejaría de ser forzosamente darwiniana. En este caso, la selección natural podría actuar básicamente sobre cambios genéticos reguladores o sobre cambios en otras fracciones del ADN.

La controversia entre neutralistas y seleccionistas sigue sin resolverse. Los defensores extremos del neutralismo ven el mantenimiento de la variabilidad genética como una acción conservadora de la selección frente a las variaciones caóticas del azar (mutación y deriva genética, básicamente). Conciben el mundo biológico como un modelo de equilibrio frente al cual se alzan constantemente amenazas perturbadoras. Por el contrario, los apóstoles extremos del seleccionismo ven la variedad inherente a los seres vivos como su propia realidad y conciben la selección como una fuerza que favorece la presencia de varios tipos o formas en vez de una sola. Para un seleccionista, el polimorfismo no es el resultado de la lucha para mantener un "statu quo", sino el único status renovador y capaz de futura evolución. Por esto, el heterocigoto "debe" ser superior.

La teoría darwiniana de la selección natural sigue siendo la más coherente para explicar el fenómeno evolutivo, pero la presente controversia está aportando nuevos puntos de vista que pueden ampliar favorablemente nuestra idea de la evolución. Es muy probable que, como ha sucedido en otras controversias históricas, el problema sea de enfoque. Ni los neutralistas niegan que la selección natural actúe en la evolución, ni los seleccionistas proscriben el papel de la deriva genética en la misma. Muchos evolucionistas están actualmente tratando, no ya de medir la variabilidad genética, sino de detectar qué variabilidad genética es la importante para el cambio evolutivo. La moderna genética molecular nos ha permitido detectar e interpretar la variabilidad proteica, pero esto ha conducido a la búsqueda de nuevos tipos de variabilidad genética quizás más directamente relacionados con cambios evolutivos en los momentos cruciales de la especiación. El creciente interés que la controversia neutralismo-seleccionismo ha suscitado evidencia que ahora más que nunca la genética evolutiva vuelve a ser un centro de atracción en biología. Este interés trasciende el ámbito del especialista científico y abarca a todos los hombres preocupados por su posición como seres vivos en este planeta. Dobzhansky, uno de los fundadores más cualificados de la moderna teoría evolutiva, ha resumido magistralmente este sentido humanista de la evolución biológica diciendo que "nada tiene sentido si no se contempla a la luz de la Evolución".

Juegos matemáticos

Donde el Conde Drácula, Alicia, Porcia y otros muchos participan en diversos rompecabezas lógicos

Martin Gardner

"Les presento ahora al Profesor Smullyan, quien les demostrará que o bien él no existe o bien ustedes no existen; pero ustedes no sabrán cuál es el caso."

–MELVIN FITTING, presentando a Raymond Smullyan en un club de estudiantes de matemáticas

El libro de Raymond M. Smullyan, titulado *What Is the Name of This Book?* (¿Cómo se llama este libro?), cuya aparición está prevista para este mes de mayo, editado por Prentice Hall, Inc., es la más original, más profunda y más humorística colección de problemas de lógica recreativa que se haya escrito jamás. Contiene más de 200 problemas totalmente nuevos, tramados todos ellos por este ingenioso autor, sazonados con bromas matemáticas, animadas anécdotas y paradojas capaces de volver loco al más pintado. El libro culmina con una serie de problemas, presentados como narraciones, que conducen al lector hasta el núcleo del revolucionario trabajo que el fallecido Kurt Gödel realizó sobre la noción de indecidibilidad.

¿Quién es Smullyan? Nació en 1919, en Nueva York, estudió filosofía con Rudolf Carnap en la Universidad de Chicago, y se recibió de doctor en matemáticas por la Princeton University. En la actualidad es catedrático de matemáticas en el Lehman College de la City University de Nueva York, y es miembro del claustro de los departamentos de matemáticas y de filosofía del City University Graduate Center. Entre los especialistas de lógica, de teoría de recursión, teoría de demostración y de inteligencia artificial, Smullyan es mucho más conocido como autor de dos pequeños y elegantes tratados: *First-Order Logic* (Springer-Verlag, 1968) y *Theory of Formal Systems* (Princeton University) Press, 1961). Su artículo sobre el famoso problema del continuo de Georg Cantor, en *The Encyclopedia of Philosophy,* es una maravilla por su densidad y lucidez. El año pasado Harper & Row publicó su primer trabajo no perteneciente a su especialidad, *The Tao is Silent,* una de las mejores introducciones al taoísmo que haya tenido ocasión de ver.

Las principales aficiones de Smullyan son la música (es un consumado pianista de clásica), la magia (en su juventud actuó como profesional, aunque no como ocupación exclusiva) y el ajedrez. Algunos de sus brillantes problemas de ajedrez han sido publicados en esta sección. Recientemente ha terminado dos colecciones de problemas de ajedrez, *Chess Mysteries of Sherlock Holmes* y *Chess Mysteries of the Arabian Knights,* en las que cada problema es adecuadamente ambientado. Añádase un seductor estilo literario, un gran sentido del humor y el gusto por lo paradójico de L. Carroll, y se podrá captar el sabor de *What Is the Name of This Book?*

El libro comienza con una historia verídica que presenta ya uno de los temas constantes en Smullyan. Un Día de los Inocentes, cuando Smullyan tenía seis años, su hermano mayor, Emile, le dijo que le iban a tomar el pelo como nunca. Todo el día estuvo Smullyan esperando la mala pasada; todavía esperándola, permaneció despierto aquella noche. Finalmente, Emile le reveló la burla: Raymond había estado esperando ser burlado, y de esta forma, sin hacer nada, Emile se había burlado de él.

"Me acuerdo que le estuve dando vueltas en la cama hasta mucho después de apagar la luz", escribe Smullyan, "sin estar seguro de que verdaderamente me hubieran tomado el pelo. Por una parte, si yo no había sido burlado entonces yo no había tenido lo que esperaba, y, por consiguiente, me había engañado... Pero con igual razón podía decirse que si me había tomado el pelo entonces yo había sufrido lo que esperaba, así que ¿de qué forma me habían engañado?... No resolveré ahora esta paradoja; retornaremos a ella, presentada de una forma u otra, varias veces en el discurso del libro."

Tras una sección introductoria en la que se exponen algunos problemas clásicos, presentados muchos de ellos con nuevas y divertidas variantes, Smullyan introduce tres tipos de personajes que intervendrán en la mayoría de los problemas subsiguientes: los "caballeros", quienes siempre dicen la verdad; los "pícaros", quienes mienten siempre; y los "normales", quienes mienten unas veces y otras dicen la verdad.

Es sorprendente lo que puede deducirse de unas cuantas líneas de diálogo con tales personajes. Por ejemplo, en una isla habitada solamente por caballeros y pícaros, Smullyan se encuentra a dos hombres descansando al pie de un árbol. "¿Es caballero alguno de ustedes dos?", les pregunta. Cuando uno de ellos responde –llamemos por ejemplo *A* a este hombre– Smullyan conoce instantáneamente la solución de su dilema. ¿Es *A* caballero o pícaro? ¿Qué es el otro hombre?

Parecería que no hay suficiente información para resolver este problema, pero la clave reside en el hecho de ser la respuesta de *A* suficiente para que Smullyan hallase la solución. Si la respuesta hubiese sido "Sí", Smullyan no hubiese obtenido información. (Si *A* fuese caballero, podría ser caballero uno o ambos de los dos; si *A* fuese pícaro, ambos podrían ser pícaros.) Por consiguiente, *A* tuvo que haber contestado "No". Ahora bien, si *A* fuese caballero, hubiera tenido que contestar "Sí", y puesto que ha dicho "No", forzosamente ha de ser pícaro. Como el que habla es pícaro, su manifestación debe ser falsa; así pues, hay al menos un caballero presente. Por tanto, el otro hombre es caballero.

Pronto hace acto de presencia Alicia, el personaje de Lewis Carroll. Nos la encontramos vagabundeando por el Bosque del Olvido, donde es incapaz de recordar el día de la semana (véase el capítulo 3 de *Alicia a través del espejo*). En el bosque, Alicia se encuentra con el León y el Unicornio. El León miente los lunes, martes y miércoles, y el Unicornio miente los jueves, viernes y sábados. En todas las demás ocasiones, ambos animales dicen la verdad. "Ayer me tocó mentir", dice el León. "También ayer me tocó mentir", dice el Unicornio. Alicia, que es tan despierta como Smullyan, es capaz de deducir en qué día

de la semana se encuentra. ¿De qué día se trata? Daré el próximo mes la solución de este problema y de los siguientes.

Aparecen más personaje de *A través del espejo:* los hermanos Tweddle, el Rey Blanco, Humpty Dumpty y el Jabberwock. Tweedledee se conduce como el Unicornio, y Tweedledum, como el León. Tras resolver Alicia cierto número de problemas basados en su conversación con los Tweedles, Humpty Dumpty revela un secreto cuidadosamente guardado del *Espejo:* hay un tercer hermano, de aspecto idéntico al de Dum y Dee, llamado Tweedledoo. Doo siempre miente. Este descubrimiento hace que Alicia se enfade muchísimo, porque todas sus deducciones precedentes pueden ahora ser falsas. Por otra parte, Humpty puede estar mintiendo, y Tweedledoo puede no existir. Se dan cuatro descripciones de lo que a continuación ocurre, y se le pide al lector que deduzca cuál es la verdadera, y si Tweedledoo existe o no.

La acción se desplaza después a *El mercader de Venecia,* de Shakespeare, en la célebre ocasión en que Porcia presenta a su pretendiente tres arquillas, de oro, plata y plomo, respectivamente, cada una de las cuales lleva grabada una inscripción diferente. Solamente una de las arquetas contiene el retrato de Porcia, y si el pretendiente la elige correctamente, Porcia habrá de casarse con él. (Incidentalmente, no suele observarse que Porcia le da con gran sutileza una preciosa indicación a su pretendiente cuando canta: "Tell me where is fancy bred/Or in the heart or in the head?" Las palabras finales de ambos versos riman con "lead" (plomo), que es la elección correcta.)

Cambiando las inscripciones de las arquillas, Smullyan crea una serie de notables problemas que van acercando más y más al lector al descubrimiento de Gödel. Se muestra el primero de estos problemas en la ilustración inferior. Porcia, quien jamás miente, explica a su pretendiente que a lo sumo una inscripción es verdadera. ¿Qué cofrecillo deberá elegir?

Smullyan improvisa agudas variaciones en torno al tema de Porcia. En algunos problemas cada cofre porta dos inscripciones. También se nos informa que hay dos artesanos fabricantes de arquillas: Bellini, que siempre pone en sus cofres inscripciones verdaderas, y Cellini, que siempre las pone falsas.

Examinemos otro problema de este tipo, que se muestra en la ilustración de la página siguiente. El cofrecillo de oro porta la inscripción "El retrato no se encuentra aquí", y el de plata, la leyenda "Exactamente uno de estos dos enunciados es verdadero". Estas dos inscripciones ofrecen un dilema lógico de enorme importancia en la historia de la semántica moderna. El pretendiente razona así: "Si el enunciado de plata es verdadero, entonces el de oro es falso. Si el enunciado de plata es falso, entonces las inscripciones de ambos cofres son, o bien ambas verdaderas o bien ambas falsas. Si la inscripción en plata es falsa, no pueden ser ambas verdaderas, y, por tanto, son ambas falsas. En cualquiera de los casos posibles el enunciado escrito en oro es falso; por tanto, el cofrecillo de oro ha de contener el retrato. Con aire triunfal, el pretendiente abre la arquilla de oro, descubriendo con horror que está vacío. El retrato se encuentra en el otro cofrecillo. ¿En qué falla su razonamiento?

El error consiste en suponer que el enunciado del cofrecillo de plata es o bien verdadero o bien falso. El problema nos enreda con la moderna noción de metalenguaje. La discusión de los valores de verdad de un lenguaje no es permisible sino en el seno de un lenguaje mayor, metalenguaje del primero, al que contiene como subconjunto de su conjunto de términos. Cuando un lenguaje hace referencia a sus propios valores de verdad resultan con frecuencia contradicciones lógicas. Sin disponer de metaenunciados relativos a la veracidad o falsedad de las inscripciones de los cofrecillos, o de información acerca de las interrelaciones entre sus valores de verdad, las inscripciones de los cofres pueden resultar carentes de significado.

Caminando con su larga zancada, entra ahora en el libro el Inspector Leslie Craig, de Scotland Yard. Smullyan nos propone una variada serie de enigmas extraídos de los archivos del Inspector, cuyo esclarecimiento es posible mediante una deducción lógica cuidadosa. El primero es el más sencillo de todos ellos:

"Se ha robado una gran cantidad de mercancías de un almacén. El ladrón (o ladrones) se ha llevado el botín en un coche. Tres conocidos delincuentes, *A*, *B* y *C*, son detenidos y llevados a Scotland Yard para ser sometidos a interrogatorio. Se han podido establecer con certeza los siguientes hechos:

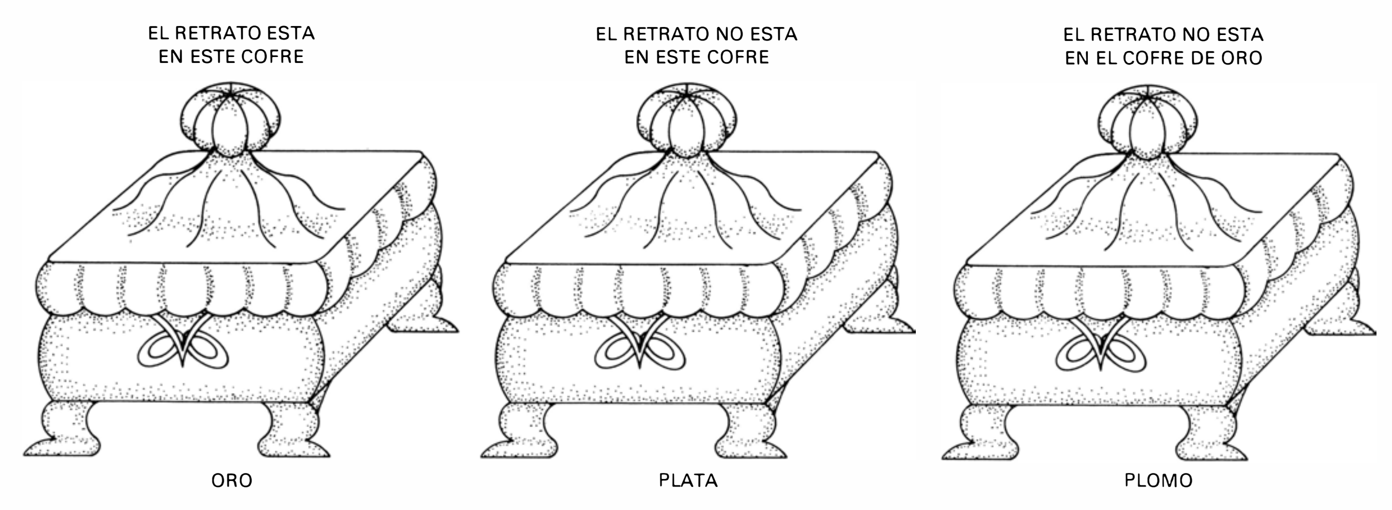
"(1) Aparte de *A*, *B* y *C*, nadie más ha intervenido en el robo.

"(2) *C* nunca 'trabaja' sin la ayuda y complicidad de *A* (y, posiblemente, de terceros).

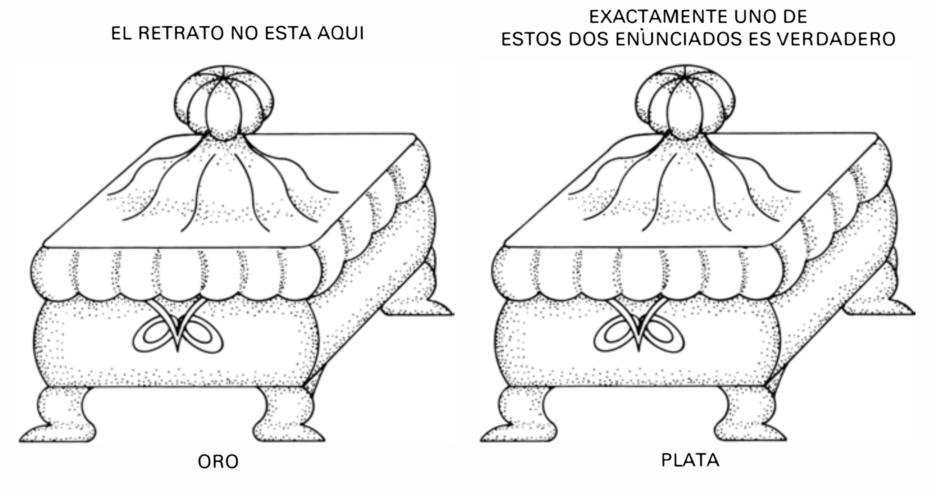
"(3) *B* no sabe conducir.

"¿Es *A* inocente o culpable?".

En las páginas siguientes, Smullyan se interesa por temas de tanta importancia práctica como el modo de evitar encuentros con los hombres-lobos, de elegir esposa, de defenderse a sí mismo ante un tribunal o de casarse con la hija de un rey. Por ejemplo, imaginemos que usted desea convencer a una joven casadera, quien reserva sus afectos para pícaros ricos, de que usted es uno de ellos. (Se



La primera prueba de Porcia, extraida de **¿Cómo se llama este libro?**



¿Dónde está el retrato de Porcia?

supone que usted es rico o es pobre.) ¿Es posible conseguir convencerla con sólo una frase? Pues sí. Basta que usted diga: "Soy un pícaro pobre". La joven sabría inmediatamente que usted no es "caballero", porque un caballero no podría mentir diciendo que es un pícaro pobre. Puesto que usted es un pícaro, su afirmación tiene que ser falsa; así pues, usted tiene que ser un pícaro rico. Supongamos ahora que la muchacha solamente se sienta inclinada hacia los caballeros. ¿Con qué frase podría convencerla de que usted es un rico caballero?

La sección siguiente presenta problemas lógicos basados en enunciados condicionales de la forma: "Si P es verdadero, entonces Q es verdadero". Los dos enunciados están conectados por la relación de implicación, cuya comprensión es absolutamente indispensable para entender el cálculo proposicional. Smullyan juguetea con las paradojas más familiares que presenta la noción de implicación, ofreciendo después 18 ingeniosos problemas que ningún lector podrá elucidar completamente sin aprehender firmemente los principios lógicos necesarios.

El siguiente cuadro se desarrolla en la Isla de Baal, lugar único en el mundo, pues en él hay alguien que conoce la respuesta de la superúltima cuestión de la metafísica: ¿Por qué hay algo que existe? La isla está poblada solamente por "caballeros" y por "pícaros". Tras una serie de encuentros con los lugareños, Smullyan demuestra que la Isla de Baal no puede existir.

Tal demostración de inexistencia no es válida para la siguiente de las islas que Smullyan visita: la isla de los Zombis. No hay forma de distinguir a los zombis, que siempre mienten, de los seres humanos, que siempre dicen la verdad. La vida se complica todavía más por el hecho de que todas las preguntas que normalmente se contestarían "sí" o "no" se responden aquí con "bal" o "da", pero no se sabe cuál significa "sí" y cuál significa "no". Supongamos que se le pregunta a un lugareño si "bal" significa "sí", y que responda "bal". Es imposible saber lo que significa "bal", pero ¿podría saberse si es un humano o un zombi? ¿Es posible descubrir lo que significa "bal" planteando una sola pregunta cuya respuesta sea "bal" o "da"?

Transilvania no es menos desconcertante. En ella, el aspecto de los seres humanos (que dicen la verdad) es idéntico al de los vampiros (todos ellos mentirosos), y la mitad de los habitantes están locos. Los locos creen que todas las proposiciones verdaderas son falsas y que todas las proposiciones falsas son verdaderas. Así pues, hay cuatro tipos de transilvanios: seres humanos cuerdos, seres humanos locos, vampiros cuerdos y vampiros locos. Evidentemente, todo lo que diga un humano cuerdo es verdadero, y todo lo que diga un humano loco es falso. Recíprocamente, cualquier cosa que diga un vampiro cuerdo es falso y todo lo que diga un vampiro loco es verdadero. Afortunadamente, las preguntas se responden en español. ¿De qué forma podría determinarse con una sola pregunta si un transilvanio es o no un vampiro? ¿Sabría el lector determinar con una sola pregunta si se trata de un loco?

Smullyan, ansioso por saber si el Conde Drácula está vivo o muerto, le hace la pregunta a varios transilvanios. Se le pide al lector que deduzca la solución a partir del diálogo. Esta sección culmina con un baile de gala en el castillo del Conde Drácula, debido a que todas las preguntas son contestadas "bal" o "da", como en la isla de los zombis. Hay en consecuencia tres variables a tener en cuenta: El que habla ¿está cuerdo o loco? ¿Es humano? ¿Qué significa "bal"? Finalmente, Smullyan descubre que Drácula vive, aunque está loco.

Uno de los capítulos del libro se titula "Cómo demostrar una cosa cualquiera". Tras examinar un sofisma del *Eutidemo,* uno de los diálogos de Platón, en el cual uno de los oradores demuestra que el padre de su antagonista es un perro, Smullyan discute cierto número de curiosos artificios mediante los cuales se puede demostrar la existencia de prácticamente cualquier cosa –de Dios, de Satanás, de los unicornios, de Papá Noel, etcétera–. Uno de estos artificios ha sido deducido del tradicional argumento ontológico de la existencia de Dios. Algunos de los restantes son variantes de un sutil método descubierto por el lógico-matemático J. Barkley Rosser.

Tomemos, por ejemplo, la siguiente oración: "Si esta oración es verdadera, entonces existe Papá Noel". Smullyan escribe en su libro: "Si la oración es verdadera, entonces ciertamente existe Papá Noel (porque si la oración es verdadera, entonces también ha de ser cierto que si la oración es verdadera entonces Papá Noel existe, de donde se deduce que Papá Noel existe); por consiguiente, se da el caso que dice la oración, y así pues, la oración es verdadera. Por tanto, la oración es verdadera, y si la oración es verdadera entonces existe Papá Noel. De aquí se sigue que Papá Noel existe". El razonamiento es falaz, pero sin una buena comprensión del papel de los metalenguajes no es fácil explicar exactamente por qué.

El penúltimo capítulo presenta la conocida "paradoja del mentiroso" ("Este enunciado es falso"), y muchos de sus disfraces y variantes. Smullyan presenta algunas de las más profundas paradojas de la lógica y de la teoría de conjuntos con una claridad que nunca se había conseguido hasta ahora. Por ejemplo, he aquí su explicación de la famosa paradoja, conocida como paradoja de Richard, en que se basa la demostración del teorema de indecidibilidad de Gödel.

Un matemático tiene un libro titulado *El libro de los conjuntos*. En cada una de sus páginas está definido un conjunto de números naturales, ya sea dando la lista de los números que lo componen, ya mediante una descripción de los números que pertenecen a él. Las páginas del libro están numeradas consecutivamente. ¿Podremos describir un conjunto de enteros positivos que sea imposible encontrar en el libro?

Podemos, en efecto. Si un número natural n pertenece al conjunto definido en

la página *n*, lo llamaremos número extraordinario. Si *n* no pertenece al conjunto de la página *n*, diremos que es un número ordinario. Examinemos el conjunto de todos los números ordinarios. Supongamos que este conjunto estuviese definido en una cierta página. El número de esa página no podría ser ordinario, pues, de serlo, tal número pertenecería al conjunto definido en esa página, y, por consiguiente, sería un número extraordinario. Por otra parte, tampoco puede ser número extraordinario, pues en tal caso tendría que figurar entre los del conjunto de esa página, y supusimos que tal conjunto solamente está formado por números ordinarios. Esta contradicción nos obliga a abandonar la hipótesis de que el conjunto de números ordinarios pueda figurar entre los inventariados en el libro.

Nos encontramos ahora preparados para el apoteósico capítulo que Smullyan dedica al descubrimiento de Gödel. Este capítulo final es la mejor introducción que conozco para comprender la gran línea divisoria que en el estudio de los fundamentos de la matemática ha supuesto tal descubrimiento. Desde los tiempos de Leibniz han soñado los matemáticos que algún día llegarían todas las matemáticas a unificarse en un vasto sistema donde pudiera demostrarse la veracidad o falsedad de todo enunciado correctamente construido. Leibniz extendió este ideal incluso a las disputas filosóficas. "En el caso de que surgieran controversias", escribió, "tan poco necesario sería que disputasen los filósofos como que lo hicieran los contables. Pues bastaría con que tomando un lápiz cada uno, y sentándose con su pizarra, se dijeran uno a otro (con un amigo por testigo, si lo deseasen): Calculemos".

Tal ideal quedó hecho añicos para siempre al aparecer en 1931 el artículo de Gödel. Gödel, que contaba entonces 25 años, demostró en su trabajo que los *Principia Mathematica*, el sistema deductivo de Alfred North Whitehead y Bertrand Russell (y lo mismo vale con otros sistemas relacionados con aquél, como la teoría de conjuntos ordinaria), contienen proposiciones indecidibles, es decir, proposiciones que son ciertas, pero cuya veracidad no puede ser establecida dentro del sistema. Con mayor precisión, Gödel demostró que si un sistema como el de los *Principia Mathematica* satisface ciertas condiciones razonables, como la de ser coherente (libre de contradicciones) entonces permite construir proposiciones que son indecibiles. Al mismo tiempo, demostró que si un tal sistema es coherente, no existe forma de demostrar la coherencia del sistema dentro del sistema.

Los resultados anteriores son válidos para todo sistema deductivo lo suficientemente rico y complejo como para contener la aritmética. Incluso en la aritmética ordinaria hay proposiciones que son verdaderas, pero indemostrables. (Hay sistemas muy simples, como la aritmética sin multiplicación, que se hallan libres de enunciados indecidibles.) No es posible, además, demostrar la coherencia de la aritmética por medios aritméticos solamente.

Desde luego, se puede ampliar la aritmética añadiendo nuevos axiomas, de forma que en el sistema así ampliado sea posible demostrar cualquiera de los enunciados anteriormente indecidibles. Desdichadamente, la situación no es por ello menos desesperada que antes: el mismo razonamiento de Gödel permite demostrar que ahora el sistema ampliado contiene nuevos enunciados indecidibles, y que tampoco podrá establecerse la coherencia del sistema razonando en el seno del sistema. La construcción de sistemas cada vez mayores puede proseguir indefinidamente, pero nunca se alcanzará un nivel donde se hayan expurgado todas las proposiciones indecidibles, o en el que pueda prepararse una demostración de coherencia del sistema que utilice solamente elementos del sistema.

Hay en aritmética un famoso problema no resuelto, llamado conjetura de Goldbach, que afirma que todo número par mayor que dos es suma de dos números primos. Nadie ha podido ni demostrarla ni hallar un contraejemplo. Es posible que la conjetura de Goldbach sea un indecidible de Gödel. De ocurrir así, ello significa que la conjetura es verdadera, pero indemostrable dentro de la aritmética. Sería verdadera, porque si fuese falsa existiría un contraejemplo, y entonces la conjetura sería decidible.

La situación anterior presenta una particularidad todavía más molesta: ¡la imposibilidad de una demostración constructiva que garantice que nunca se llegará en teoría de números a encontrar una demostración de la veracidad de la conjetura de Goldbach y otra demostración de su falsedad! Los matemáticos tienen fe y esperanza en que nunca sucederá nada parecido con ningún teorema aritmético, pues de ocurrir así toda la aritmética y toda la matemática superior quedaría reducida a escombros. (Es fácil demostrar que si un sistema deductivo contiene al menos una contradicción, entonces es posible demostrar cualquier proposición contenida en el sistema.) Los matemáticos de escuela platónica, que consideran que los axiomas de la aritmética son verdaderos y que las reglas de razonamiento e inferencia son correctas, no sufren tales angustias, pues están convencidos de que no pueden surgir contradicciones. Los matemáticos de la escuela formalista y constructivista no disponen, desde luego, de tales garantías.

Los enunciados indecidibles en el sentido de Gödel solamente son indecidibles en el seno de un sistema dado. En 1936, en trabajos publicados por A. M. Turing y Alonzo Church, quedó establecida la existencia de problemas indecidibles en un sentido más radical y profundo. Turing y Church demostraron la existencia de poblemas para los que no hay algoritmos finitos, esto es, procedimientos descritos paso por paso, que puedan resolverlos. Entre los ejemplos de tales indecidibles absolutos se encuentra el famoso problema de detención de la máquina de Turing, el problema de los dominós de color en la teoría de teselaciones, algunos problemas del juego de la vida, de John Horton Conway, y otros muchos. Nunca, por muy lejano que sea el futuro que se considere, se podrá construir en un mundo coherente desde el punto de vista lógico, un ordenador, por muy potente que sea, que manejando y combinando símbolos pueda resolver tales problemas en un número finito de pasos.

Desde 1936 hasta la fecha se han construido toda clase de demostraciones de los resultados de Gödel, así como de los de Church-Turing, algunas de ellas más simples que la original de Gödel. Smullyan presenta la demostración de Gödel de manera deliciosa, imaginando una isla de Gödel habitada solamente por caballeros y pícaros. Los caballeros que han dado pruebas de su caballerosidad se llaman caballeros probados. Los pícaros que han dado pruebas de indudable villanía se llaman pícaros probados. Los habitantes de la isla han formado "clubs" para los cuales són válidas las siguientes condiciones:

1. El conjunto de todos los caballeros probados forma un club.
2. El conjunto de todos los pícaros probados forma un club.
3. Todo club *C* tiene otro club complementario: el club formado por todos los habitantes de la isla que no son miembros del club *C*.
4. Dado un club cualquiera, existe al menos un habitante que reconoce ser miembro de dicho club.

Smullyan puede ahora demostrar en tan sólo tres párrafos de razonamiento sencillo, sin terminología técnica, que existen en la isla al menos un caballero

no probado y al menos un pícaro no probado. Si se considera que los caballeros representan enunciados verdaderos, que los caballeros probados representan enunciados demostrablemente verdaderos, que los pícaros son enunciados falsos y que los pícaros probados son enunciados demostrablemente falsos, los resultados de los razonamientos de Smullyan se corresponden con los de Gödel. Tan sólo le son precisas otras tres cláusulas para demostrar otro resultado de Alfred Tarski relacionado con los anteriores, a saber, que ni el conjunto de los caballeros ni el conjunto de los pícaros puede formar un club.

Hay una versión muy conocida de la paradoja del mentiroso en la que en lugar de una sola frase intervienen dos, escritas cada una en una cara de una ficha. Uno de los enunciados afirma: "El enunciado escrito en la otra cara de esta ficha es verdadero. El otro enunciado establece que "el enunciado escrito en la otra cara de esta ficha es falso". Ninguna de las dos frases hace mención de sí misma; no obstante, la contradicción es evidente. De forma análoga, Smullyan concibe las que denomina "islas doblemente gödelianas", pobladas por caballeros y pícaros que satisfacen la condición siguiente: Dados dos clubs C_1 y C_2, existen habitantes A y B tales que A afirma que B es miembro del club C_1 y B afirma que A lo es de C_2. El estudio de tales islas "bigödelianas" es uno de los esparcimientos favoritos de Smullyan. En su libro, Smullyan examina algunos de sus descubrimientos sobre ellas y presenta algunos problemas todavía no resueltos.

El libro termina con una versión verdaderamente asombrosa de la construcción de Gödel de una proposición no demostrable. Examinemos el siguiente enunciado: "Esta proposición nunca podrá ser demostrada". Escribe Smullyan: "Si el enunciado es falso, entonces es falso que nunca podrá ser demostrado; por consiguiente, *puede* ser demostrado, lo que significa que es verdadero. Así pues, si el enunciado es falso se tiene una contradicción, y, por tanto, debe ser verdadero.

"Ahora bien, acabo de demostrar que el enunciado es verdadero. Como el enunciado es verdadero, ocurre verdaderamente lo que afirma, a saber, que nunca podrá ser demostrado. Pero entonces, ¿cómo he podido demostrarlo?".

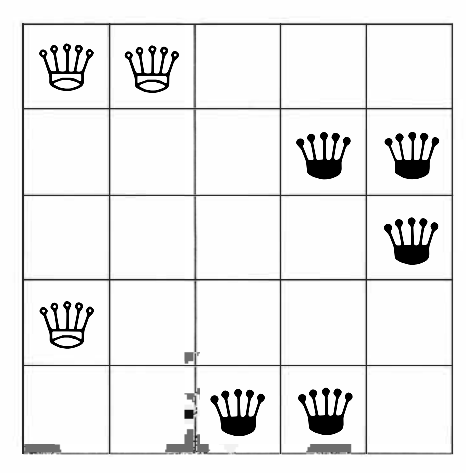
La falacia, explica Smullyan, consiste en que no se ha explicado claramente el significado de "demostrable". Examinemos una versión modificada de la paradoja: "Esta proposición no es demostrable en el sistema S". La paradoja se deshace como por encanto. "La verdad más interesante es que el enunciado anterior ha de ser una proposición verdadera que no es demostrable en el sistema S". (Se supone, evidentemente, que todo enunciado demostrable en el sistema S es verdadero.) Presentada de esta forma, la paradoja es "una tosca formulación de la proposición X de Gödel, la cual puede considerarse que afirma su propia indemostrabilidad, pero no en un sentido absoluto, sino solamente dentro del sistema dado".

Llegado a este punto Smullyan recuerda de pronto que todavía no ha contestado la pregunta "¿Cómo se llama este libro?". Su nombre es... Aunque es preferible que mire la última página del libro para conocer la respuesta.

El problema del último mes consistía en situar cinco reinas de un color y tres de otro en un tablero de ajedrez de cinco cuadros de lado, de forma que ninguna de ellas atacase a otra de distinto color. Se muestra la única solución (única salvo giros o simetrías) en la ilustración de la página opuesta.

Muchos lectores me enviaron correcciones y nuevos datos relativos a la sección sobre grafos de Ramsey del pasado mes de enero. Las acotaciones ahora establecidas para $R(4,5)$ son entre 25 y 28 y para $R(5,5)$ son 42 y 55. La cota inferior 42 fue obtenida por Yi-Hsin Chen, para su reciente tesina de licenciatura en la Universidad de Toronto. Gary Lorden corrigió mi afirmación de que Léopold Sauvé había demostrado que todo grafo extremal puede ser bicoloreado de manera que todos los triángulos monocromáticos sean de igual color. A.W. Goodman había demostrado con anterioridad que el enunciado es verdadero para grafos completos de número par de vértices. Sauvé demostró que es falso para grafos de un número impar de puntos, con la única excepción de K_7, descubierto por Paul Erdös. El artículo de Lorden sobre "Blue-Empty Chromatic Graphs" *(The American Mathematical Monthly,* vol. 69, n.º 2, págs. 114-120: febrero de 1962) da métodos sencillos de construcción de los coloreados. Thomas Hoover señaló cómo pueden usarse matrices de adyacencia para hallar tales coloreados.

Robert E. Greenwood y Frank Harary me han proporcionado la temprana historia del famoso problema de demostrar que entre seis personas cualesquiera hay siempre al menos tres que se conocen y al menos tres que se desconocen mutuamente. Hacia 1950 el problema rodaba ya de mano en mano, y por sugerencia de Harary fue incluido en la William Lowell Putnam Mathematical Competition de 1953.



Solución del problema de ajedrez del mes pasado

Alan Cassel, del Instituto de Tecnología de Massachusetts, fue el primero en señalar que el Dr. Matrix había cometido un descuido y se había equivocado al afirmar la existencia de un gran número primo de dígitos consecutivos. El número correcto es 1234567891234567891234567891. Utilizando un nuevo y rápido método de Montecarlo para comprobar el carácter primo de un número, Cassel hizo el notable descubrimiento de que repitiendo siete veces el ciclo 123456789 y añadiéndole a continuación el grupo 1234567 resulta un número pseudo-primo cuya probabilidad de ser primo verdadero es del orden de un billón a uno. Quizás alguno de los lectores pueda demostrar que es efectivamente primo.

Stanley J. Benkoski escribió diciéndome que mi teoría de la posibilidad de que las cucarachas desencadenasen terremotos por tele-psico-kinesis estaba adquiriendo credibilidad en California. "El Gobernador Brown ha pedido que todos los funcionarios de ese Estado tengan diariamente media hora de permiso en sus destinos para contribuir a la erradicación de las cucarachas. Según estimaciones de fuente oficial, la total extinción de las cucarachas en el estado de California se producirá el 13 de julio de 1978, y a partir de esa fecha los terremotos serán relegados a meras curiosidades históricas."

[Miguel Angel Lerma, de Madrid, y Borja P. Oñate y Javier Suarez, de Zaragoza, nos han señalado un error en la sección "Juegos matemáticos" del número de febrero pasado. Al ordenar alfabéticamente las expresiones nominales de los números de 0 a 1000, el último lugar lo ocupa el "21", y no el "1", como equivocadamente se dijo. De la misma forma, el número inmediatamente precedente es "23", y no "333" como se publicó en el apartado de soluciones del mes de marzo"].

Taller y laboratorio

Ilusiones ópticas que pueden producirse al colocar un filtro obscuro delante de uno de nuestros ojos

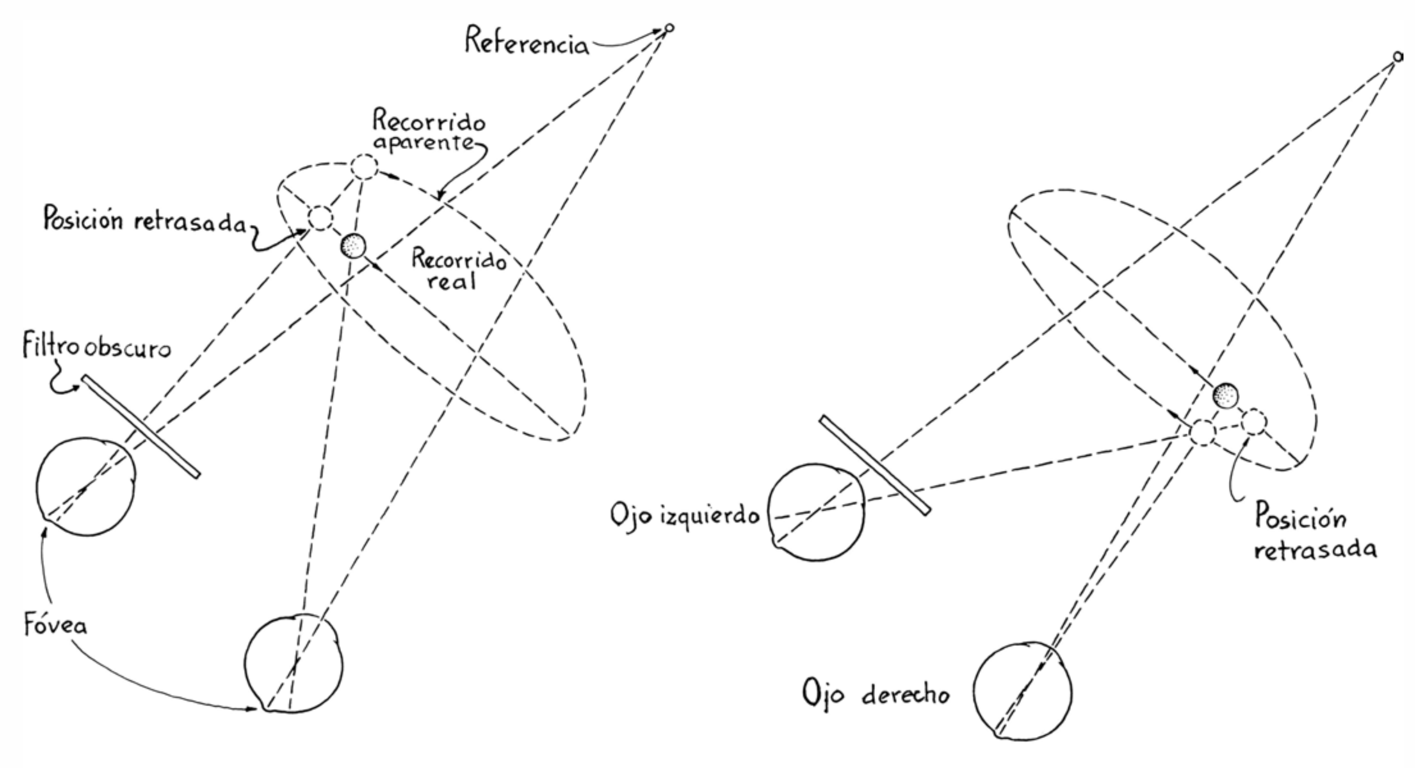
Jearl Walker

Bajo determinadas circunstancias, el movimiento oscilatorio de un péndulo que se mueva en un plano perpendicular a la línea de visión del observador puede dar la impresión de que dicho péndulo se mueve alrededor de un eje vertical central describiendo una elipse. El efecto fue explicado por primera vez en 1922 por el físico alemán Carl Pulfrich, y en su honor se reconoce desde entonces como "efecto Pulfrich". El lector puede hacer, sin dificultad, una demostración práctica de dicho efecto.

Para ello sólo necesita tres cosas: un péndulo simple, consistente en un peso colgado de una cuerda, un filtro óptico parcialmente obscurecido, como por ejemplo uno de los cristales de unas gafas de sol, y un objeto vertical, delgado. Haga oscilar el péndulo en un plano perpendicular al de su línea de visión, y sitúe el objeto alargado detrás del péndulo y en posición vertical. El filtro obscuro se colocará delante de uno de los ojos y se mirará con ambos al objeto del fondo. Parecerá que el péndulo se mueve elípticamente. Con el cristal enfrente del ojo izquierdo la rotación será en el sentido de las agujas del reloj visto desde arriba; con el filtro obscuro enfrente del ojo derecho, el movimiento será contrario al de las agujas del reloj. Cuanto más obscuro sea el filtro mayor será la profundidad aparente del recorrido elíptico del péndulo. De alguna manera el cristal obscuro provoca que el observador perciba un movimiento en profundidad, aunque no exista dicha profundidad.

Pulfrich explicó el efecto a raíz de una sugerencia que le hizo un compañero en el sentido de que tal vez el filtro obscuro retrasaba la velocidad con que la imagen que el ojo afectado recibía llegaba al cerebro. (Pulfrich era tuerto, y nunca podría haber visto tal ilusión.) La sugerencia del compañero resultó ser cierta. Cuando se disminuye la iluminación que recibe uno de nuestros ojos, la percepción de los objetos por dicho ojo está retrasada con respecto a la percepción del otro ojo en varios milisegundos (unas cuantas decenas).

Supongamos que el ojo izquierdo está tapado por un cristal obscuro y que el péndulo comienza a oscilar de izquierda a derecha. Los ojos estarán fijos en el objeto del fondo, de manera que la imagen del mismo caerá sobre la fóvea (pequeña depresión existente en la retina, que es por donde la visión es más aguda). En los comienzos de la oscilación, es decir, con el péndulo en su posición a la izquierda, la imagen del peso se formará fuera del centro de la fóvea, pero debido al retraso visual causado por el ojo izquierdo, la imagen del objeto en dicho ojo se formará bastante más desplazada todavía del centro de la fóvea. Por el ojo izquierdo, el observador verá el objeto donde estaba varias decenas de milisegundos antes, esto es, más a la izquierda de su verdadera posición. La imagen se forma en la retina más alejada del centro de la fóvea de lo que se hubiera formado si no estuviera el cristal. Lo que el observador percibe realmente es el ángulo entre el peso del péndulo y el objeto de referencia del fondo. El ángulo para el ojo izquierdo es mayor que para el ojo derecho y eso le hace percibir una diferencia que el observador interpreta como si el peso estuviera más



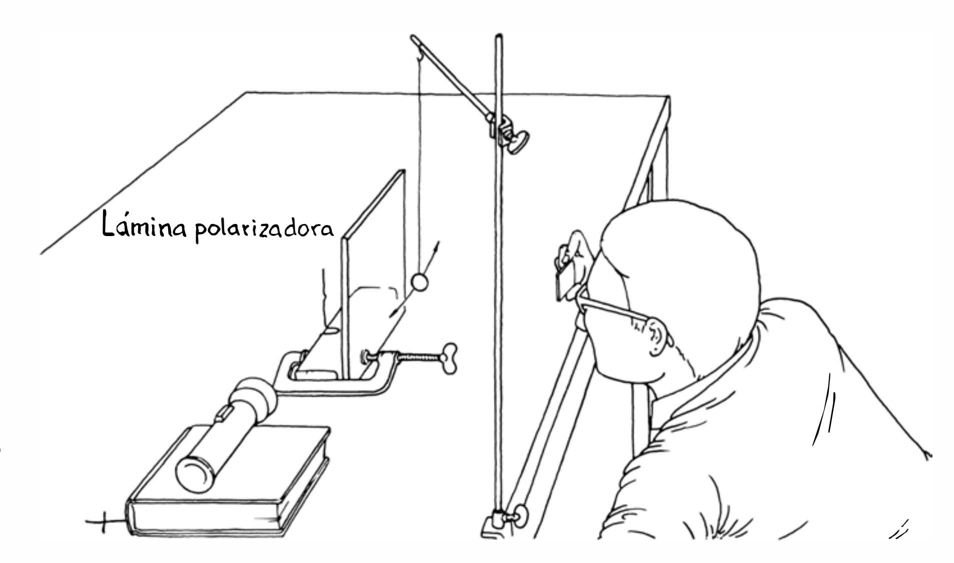
Movimiento aparente del peso del péndulo en el efecto Pulfrich

lejos de lo que realmente está. Por consiguiente, el retraso en la recepción de la imagen le da la impresión de profundidad en el movimiento del peso del péndulo.

Como la velocidad del peso se hace mayor conforme se aproxima a su centro de oscilación, la diferencia angular entre el peso y el objeto de referencia será mayor y, por tanto, la aparente profundidad del peso aumentará. Al contrario, como la velocidad del péndulo se hace menor conforme se aproxima al extremo derecho de la oscilación, la diferencia angular disminuirá y el peso parecerá que se mueve acercándose al observador. Durante el recorrido de regreso hacia la izquierda, el retraso en la percepción que provoca el ojo izquierdo hará que parezca que el peso se mueve en un punto más próximo al observador de lo que realmente está. Ese punto más próximo se hallará cerca del centro de oscilación del péndulo, donde la velocidad del peso es máxima. El movimiento que describe el peso en un ciclo completo es una rotación aparente a lo largo de una órbita elíptica alrededor de un eje vertical.

Al incrementar la obscuridad del filtro (su densidad óptica) aumenta el retraso en la percepción del ojo tapado y, como consecuencia, la profundidad del movimiento elíptico. Se puede controlar muy fácilmente la obscuridad del filtro usando como tal dos filtros polarizadores cruzados. Modificando el ángulo entre los sentidos de polarización de los filtros se puede seleccionar el grado de obscuridad deseado. Con suficiente obscuridad, la órbita llega a ser casi circular. Al ir intensificando la obscuridad llegará un momento en que el efecto se pierda debido a que sólo se verá el objeto a través del ojo descubierto. Entonces se percibirá el movimiento lineal como realmente es.

Los procesos físicos que provocan el retraso de la llegada de la señal óptica al cerebro no se conocen con suficiente detalle. El retraso se debe a la disminución de la intensidad de luz que llega a la retina o al cambio en la adaptación del ojo. Quizás influyan ambos factores a un tiempo. (El ojo se adapta a los cambios de la intensidad luminosa modificando el tamaño de la pupila y haciendo que una mayor o menor cantidad de pigmento intervengan en el proceso visual.) Un modelo con el que explicar el retraso en la visión puede ser un conjunto de cuatro circuitos eléctricos de retardo idénticos, cada uno de los cuales constaría de una resistencia, un condensador y un amplificador. Albert M. Prestrude, del Virginia Polytechnic Institute and State University, sugirió que las cuatro etapas podían corresponder a las cuatro capas de células (horizontal, bipolar, amacrina y ganglionar) en el recorrido visual, pero esta correlación no está del todo clara.



Montaje para conseguir la ilusión del péndulo

En la analogía, la sensibilidad del proceso visual (esto es, hasta qué punto con una pequeña señal de entrada puede producir una señal de salida suficiente) depende de los valores de la resistencia y de la ganancia de las unidades de amplificación. Aunque los valores de la resistencia no son fijos, pueden ser controlados por una señal de realimentación procedente de la señal que sale al final del sistema de visión. El tiempo de respuesta del sistema depende de los valores de la resistencia: a mayor resistencia, mayor tiempo de respuesta. El efecto combinado es que cuando el nivel de estímulo para el ojo decrece, la señal de realimentación procedente de la señal de salida en el proceso visual incrementa la resistencia y la ganancia del amplificador en cada etapa con el fin de aumentar la sensibilidad. El incremento de la resistencia, a su vez, significa que el sistema se tomará un mayor tiempo para responder. Por consiguiente, al disminuir la intensidad de iluminación en un ojo se obtiene una mayor sensibilidad y un mayor tiempo de respuesta.

Desde el trabajo de Pulfrich, el efecto ha sido estudiado en una gran variedad de aplicaciones. Algunos investigadores emplearon dos discos giratorios, uno para cada ojo. Mirando a través de un sistema de prismas y de lentes, un observador podría fundir las imágenes de los dos discos giratorios de manera que percibiera la imagen de un solo disco en rotación.

Cada disco tenía taladrada una línea radial. Si los dos discos giraban de forma sincronizada, el observador veía una sola línea. No obstante, colocando un filtro obscuro delante de uno de los ojos, se provocaba que desapareciera la sincronización de las líneas. Midiendo en cuánto había que modificar la fase entre los discos para que las líneas volvieran a aparecer sincronizadas en la rotación, el observador podía determinar el retraso que el filtro introducía.

Una técnica particularmente interesante fue utilizada por Bela Julesz, de los Bell Laboratories, y por Benjamin White, de la Universidad del estado de San Francisco. Mostraban a un observador películas de puntos aleatorios hechas con un ordenador en las cuales cada cuadro era estadísticamente independiente de los otros. A través de sistemas de visión independientes, cada ojo veía una película. En el centro de cada cuadro había puntos seleccionados al azar que formaban parte de un cuadrado.

Si sólo un ojo veía la película, la sensación de un cuadro central se perdía en el evidente maremágnum de puntos aleatoriamente distribuidos. Si cada ojo veía la película que le correspondía de manera simultánea, el cuadrado central se hacía visible y sobresalía del fondo de puntos distribuidos de forma fortuita. La sensación de estereoscopicidad era debida al ligero desfase de los puntos del ojo izquierdo en relación con los puntos vistos por el ojo derecho. La misma técnica se emplea en las películas convencionales en estereoscopio, en las cua-

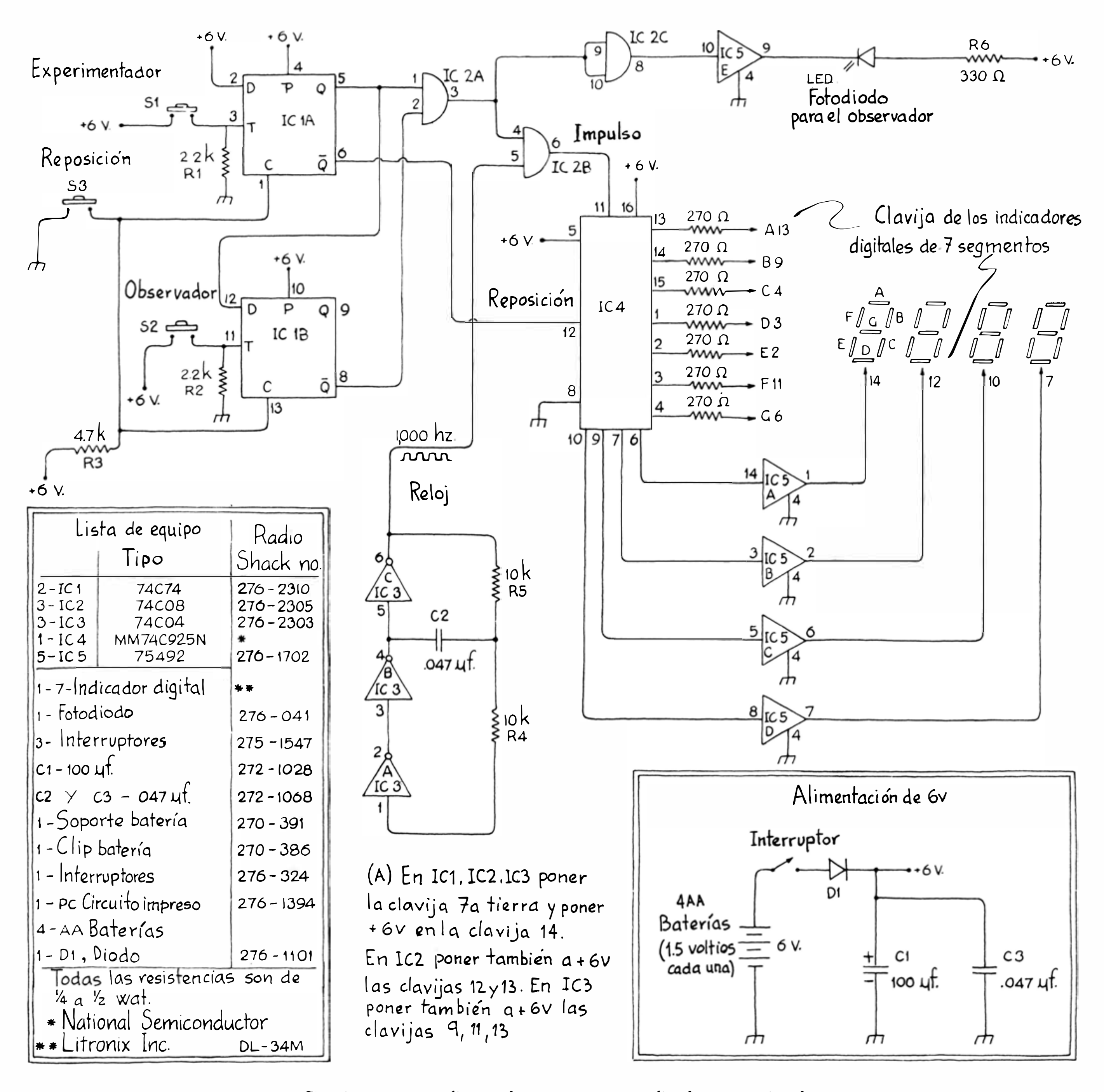
les un ojo ve la escena ligeramente distinta de la que ve el otro ojo, creando un efecto de profundidad.

En el experimento de Julesz y White, a un ojo se le mostraba, en su película, una escena más tarde que al otro ojo. Se perdía pues el cuadrado central, debido a que la selección de puntos era fortuitamente diferente de un cuadro a otro. Luego, se situaba un filtro ajustable delante del ojo que recibía la imagen adelantada. El observador ajustaba la obscuridad del filtro hasta que el retraso en la percepción del cuadro por aquel ojo era suficiente para que el cerebro recibiera, de forma simultánea, las percepciones correlativas izquierda y derecha de las escenas. El cuadrado central se hacía visible de nuevo.

En el efecto del péndulo, la profundidad aparente del movimiento pendular puede producirse por la adaptación de los ojos a la obscuridad y por el filtro obscuro. Deslumbre uno de los ojos durante unos cuantos minutos y mire entonces a un péndulo en oscilación. El efecto de profundidad se hará evidente, de la misma forma que si el otro ojo mirara a través de un filtro. El ojo deslumbrado se adaptaba a las condiciones de mayor iluminación y, por tanto, respondía más rápidamente de lo normal.

Algunos investigadores han referido que el efecto también aparece si se adapta uno de los ojos a la obscuridad. Si se tapa un ojo durante 15 o 20 minutos, y se mira luego con ambos un péndulo en oscilación, se puede percibir un esbozo del efecto Pulfrich (algunas personas lo ven, aunque yo no he podido).

Otro ejemplo de cómo las adaptaciones de nuestros sentidos nos hacen ver un mundo irreal es el hecho siguiente: Si miramos fijamente el péndulo mientras se mueve, la profundidad de su mo-



Lista de equipo Tipo		Radio Shack no.
2-IC1	74C74	276-2310
3-IC2	74C08	276-2305
3-IC3	74C04	276-2303
1-IC4	MM74C925N	*
5-IC5	75492	276-1702
1-7-Indicador digital		**
1-Fotodiodo		276-041
3-Interruptores		275-1547
C1-100 uf.		272-1028
C2 Y C3 - 047 uf.		272-1068
1-Soporte batería		270-391
1-Clip batería		270-386
1-Interruptores		276-324
1-PC Circuito impreso		276-1394
4-AA Baterías		
1-D1, Diodo		276-1101

Todas las resistencias son de ¼ a ½ wat.
* National Semiconductor
** Litronix Inc. DL-34M

Circuito correspondiente al aparato para medir el retraso visual

vimiento elíptico es función del tiempo que llevemos con el filtro obscuro situado delante de nuestro ojo. Al principio hay un retraso visual de forma inmediata y, como consecuencia, una profundidad en el recorrido del péndulo. Tanto una cosa como la otra van aumentando paulatinamente durante los siguientes veinte minutos, al cabo de los cuales el ojo cubierto se ha adaptado ya a la menor iluminación.

El movimiento elíptico del péndulo en el efecto Pulfrich normal desaparece si se mira fijamente el péndulo, en lugar de atender al fondo. Al parecer, cuando los ojos siguen al péndulo en su recorrido, en su trayectoria real, lo que logran es mantener la imagen del péndulo en la misma zona de las dos retinas, eliminando la diferencia de ángulos entre la referencia y el péndulo y, por tanto, destruyendo el efecto de la rotación. El efecto puede lograrse de nuevo aunque uno siga la trayectoria del péndulo si se obscurece el fondo de uno de los ojos. Este efecto ha sido demostrado recientemente por diferentes investigadores usando las señales de un osciloscopio en lugar de péndulos. Pero el efecto puede ser visto también con un péndulo simple.

Ponga una lámina grande de un filtro polarizador detrás del péndulo y una pequeña linterna, así como el objeto normal de referencia detrás del filtro. Delante de uno de sus ojos coloque un pequeño filtro polarizador orientado de manera que obscurezca el fondo que dicho ojo va a ver. El filtro polarizador situado detrás del péndulo debe ser lo suficientemente grande como para que el péndulo, en sus oscilaciones, no se salga de su campo. Ahora siga con la vista el recorrido del péndulo. Si el ojo izquierdo ve el trasfondo más obscuro que como lo ve el ojo derecho, parecerá que el péndulo gira en el sentido de las agujas del reloj igual que antes, a pesar del recorrido real. Aumentando la obscuridad del trasfondo mediante la rotación del filtro polarizador que habíamos colocado delante del ojo citado, se incrementa la profundidad de la trayectoria elíptica del péndulo; si se pone el pequeño filtro polarizador enfrente del otro ojo, es decir, sobre el ojo derecho, se consigue, como era de esperar, que el movimiento sea antihorario.

Como las imágenes de la lámina del filtro polarizador y la del objeto de referencia se mueven a lo largo de la retina, el incremento del retraso visual provocado por el hecho de estar la visión del ojo filtrada, da lugar al efecto de que la distancia entre el peso y el objeto de referencia esté cambiando constantemente. La mayor parte de esta profundidad variable se le atribuye al peso, tal vez porque es más fácil pensar en un cambio en la distancia que lo separa del peso que en un cambio en la distancia que lo separa del trasfondo y del objeto de referencia, ya que ambas cosas se sabe ciertamente que están quietas. Como resultado, el péndulo parece moverse a lo largo de una órbita elíptica incluso cuando uno está mirando fijamente al peso durante su recorrido.

El experimento con la ilusión del péndulo Pulfrich es una forma difícil de medir el retraso visual producido por un filtro obscuro dado. Una razón por la cual otros investigadores diseñaron experimentos distintos era el conseguir una medida del retraso con mayor precisión. En algunos experimentos realizados por mí mismo, preferí medir el tiempo de respuesta por un método relativamente simple, en el que un observador apagaba una luz que yo encendía. El tiempo se medía en milisegundos. El circuito que controlaba la luz y el reloj digital para el experimento fue diseñado y construido por el autor y por Harry Freedman, de la Universidad estatal de Cleveland.

Las pastillas de los circuitos integrados identificados en la figura como *IC* (de "integrated circuit") *1A* y *1B* eran circuitos basculantes *"D"*. Inicialmente la salida *Q* de *IC 1A* producía una señal lógica de nivel 0 y la $\bar{Q}$ una señal lógica de nivel 1. (En una señal lógica de nivel 0 no se produce diferencia de potencial; una señal lógica de nivel 1 es aquella en que el nivel de la diferencia de potencial es superior a 0.) Cuando pulsaba el interruptor *S*1, las salidas *Q* y $\bar{Q}$ de *IC 1A* cambiaban de estado; entonces *Q* producía una señal lógica de nivel 1. El estado de esta salida permanecía inalterado hasta que terminaba la prueba y yo pulsaba el interruptor de reposición *S*3, devolviendo *Q* y $\bar{Q}$ a sus estados iniciales.

Al pulsar *S*1, la señal lógica de nivel 1 de la salida *Q* de *IC 1A* iba a *IC 2A* que era una puerta "y". Esta puerta produce una salida lógica de nivel 1 únicamente si las dos entradas que recibe están en el mismo estado lógico. En este caso, *IC 1B* estaba ya enviando una señal lógica de nivel 1 a dicha puerta; cuando yo pulsaba mi interruptor, la puerta recibía dos señales de estas características y, por tanto, enviaba una señal lógica de nivel 1 a *IC 2C* e *IC 5E* para encender el diodo emisor de luz al objeto de que lo viera el observador.

Se enviaba otra señal lógica de nivel 1 desde la puerta *IC 2A* a otra puerta "y" (*IC 2B*) que la comparaba con la señal procedente de un circuito oscilador que funcionaba como reloj. A una frecuencia de 1000 veces por segundo, el reloj generaba una señal lógica de nivel 1 para *IC 2B*. Por tanto, *IC 2B* hacía una comparación 1000 veces por segundo de sus entradas y enviaba a dicha frecuencia su propia señal lógica de nivel 1 a *IC* 4. El circuito *IC* 4 era un contador de cuatro dígitos con siete segmentos de salidas decodificadas que se empleaban para activar un indicador digital de cuatro dígitos con siete segmentos, que indicaba el tiempo de respuesta del observador en milisegundos.

El *IC* 4 continúa recibiendo señales a razón de 1000 por segundo y el indicador sigue mostrando el conteo hasta que el observador pulsa el interruptor *S*2. La salida a partir de Q de *IC 1B* cambia entonces a señal lógica de nivel 0, y la puerta *IC 2A* no realiza ya por más tiempo la comparación lógica entre los *IC 1A* y *1B* basculantes. Se apaga el diodo emisor de luz del observador; la señal lógica de nivel 1 tampoco se envía más a la puerta *IC 2B*. La señal del reloj es, por tanto, ignorada y el indicador se para, mostrando el último conteo, esto es, el tiempo de respuesta del observador en milisegundos. Todo permanece así hasta que se pulse el interruptor de reposición y los dos circuitos *IC 1A* y *1B* se pongan de nuevo en sus estados iniciales.

Se debe tener cuidado de no invertir las conexiones de la batería o conectar las clavijas de suministro de energía al revés, ya que se destruirían. La resistencia *R*5 deberá ajustarse de tal forma que el reloj marque a razón de 1000 hertzios (ciclos por segundo). Esto se puede comprobar dejando que el circuito cuente durante 10 segundos como si fuera un reloj parado. Dado que la lectura más alta del contador es 9999, a los 10 segundos éste deberá registrar 0000. Entonces se debe ajustar la resistencia hasta alcanzar esta correlación. Por ejemplo, incrementando el valor de *R*5, se reduce la velocidad de conteo.

Mis observadores fueron tres estudiantes de la Universidad de Cleveland: Teresa Chandler, Sandy Hoskins y Bruce Firtha. Un observador miraba el diodo emisor de luz a través de dos grandes hojas cruzadas de filtros polarizadores. Ajustando el ángulo entre los sentidos de polarización de los filtros, yo podía seleccionar la iluminación que llegaba al observador. Para medir los ángulos del filtro fijé un poco de cinta adhesiva clara

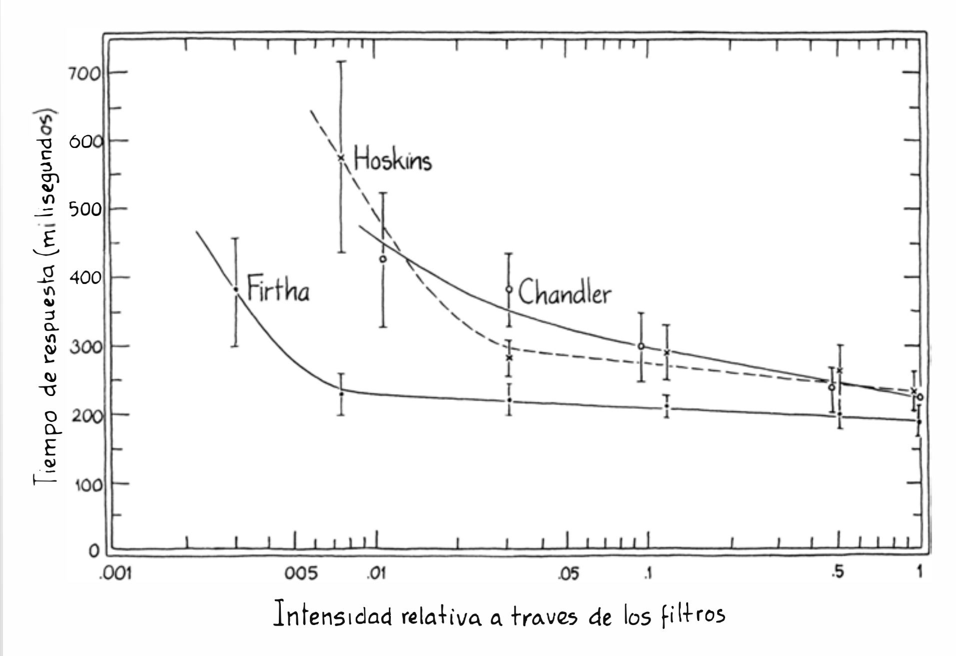
a cada filtro, de forma que las cintas fueran perpendiculares cuando los filtros lo fueran también; así bloqueaban la máxima cantidad de luz que recibían procedente de una bombilla. Se tomaron mediciones de otras orientaciones del filtro en un prolongador, colocado entre los trozos de cinta adhesiva. Probablemente, el índice de error no variaría mucho de +2 a −2 grados.

La intensidad de la luz transmitida por los filtros cruzados se comparaba con la intensidad (llamada 1) transmitida por filtros paralelos. Para cada orientación, el valor relativo de la intensidad transmitida era el cuadrado del coseno del ángulo entre los sentidos de polarización de los filtros cruzados. Estimé la intensidad relativa transmitida a través de filtros perpendiculares a partir de datos de transmisión de la Polaroid Corporation para el tipo de los filtros de polarización que estaba usando. Aunque para filtros perpendiculares la transmisión era baja, no llegaba a ser nula exactamente. Cuando se tomaron los datos con las luces de la habitación encendidas, los observadores se cubrieron la cabeza y los filtros con un paño negro, de manera que las luces no produjeran reflejos en las caras de los filtros que daban a los observadores.

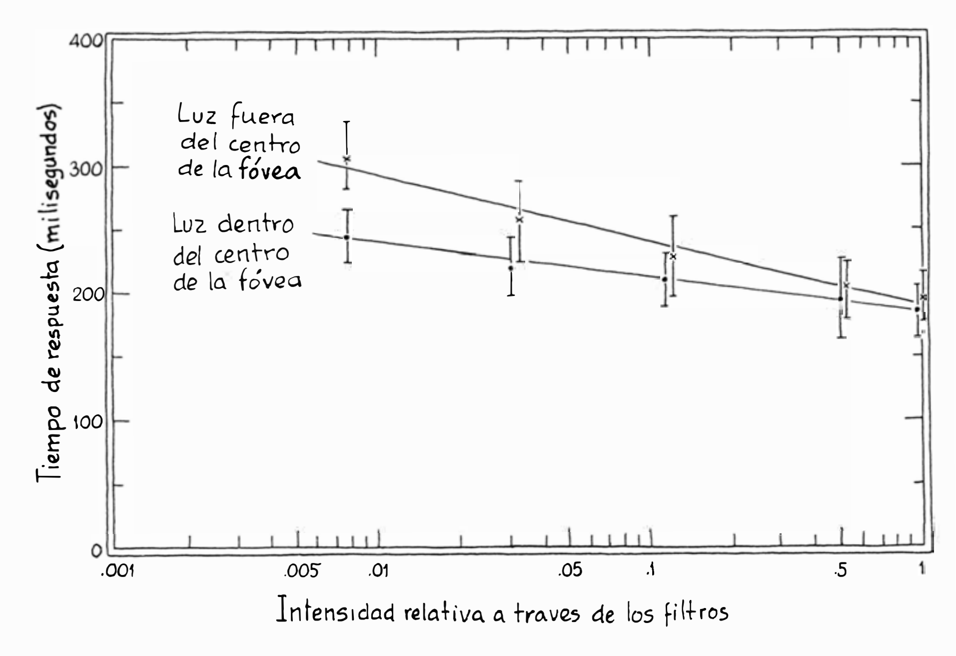
Cada vez que la luz de diodo se encendía, el observador la apagaba y yo anotaba el tiempo de respuesta. Para cada colocación de los filtros de polarización tomé, por lo menos, 50 mediciones y después estudié su significado y su desviación típica. Este tiempo medio de respuesta incluye cualquier retraso visual y también el mayor tiempo de respuesta motora del observador. Traté de encender la luz del diodo aleatoriamente, de manera que el observador no pudiera anticipar el hecho. Los tiempos de respuesta especialmente largos o cortos que resultaban ocasionalmente por el descuido o anticipación del observador se desechaban de la información.

Un problema que el experimentador puede hallar es el de encontrar un observador lo suficientemente interesado en el experimento como para concentrarse continuamente en el diodo. El experimento es cansado y la falta de interés puede aumentar considerablemente las desviaciones típicas. Mi información la tomé de observadores en una habitación obscura, dejándoles 10 minutos por lo menos para que sus ojos se adaptaran a la obscuridad. En una misma sesión se tomaron todos los datos de un observador. La información mezclada, tomada en días diferentes, podría complicar, probablemente, el experimento debido a los eventuales cambios diarios en los tiempos de respuesta motora.

En la ilustración de la página anterior puede verse la representación de parte de la información en un gráfico semilogarítmico. Si se eliminaran los tiempos de respuesta motora del experimento, los tiempos restantes de respuesta visual seguirían posiblemente este tipo de curva, pero con unos intervalos de tiempos más cortos. Es decir, el retraso visual aumentaría a medida que la iluminación del estímulo iría decreciendo; el ascenso



Representación gráfica del tiempo de respuesta de tres observadores a medida que se iba incrementando la intensidad de la luz



Respuestas cuando la luz de estímulo coincidía con el centro de la fóvea o fuera de ella

sería mayor con niveles de iluminación bajos.

Recogí datos también con las luces de la habitación encendidas. Algunos investigadores han descubierto que el retraso visual decrece cuando aumenta la iluminación total que rodea al estímulo. Sin embargo, dentro de las desviaciones típicas de mis datos, no he encontrado diferencia significativa en la curva seguida por respuesta de un observador con las luces de la habitación encendidas y con ellas apagadas.

Cuando uno de mis observadores miraba el diodo, la imagen del mismo se formaba en la fóvea, que posee conos densamente empaquetados, pero carece de bastones. Algunas investigaciones publicadas indican que el retraso visual puede ser diferente si la imagen del estímulo cae en cualquier otro lugar (es decir, fuera de la fóvea), en parte porque la respuesta visual envolvería entonces a los bastones. El lector puede tratar de conseguir este efecto haciendo que el observador mire a un objetivo pequeño desplazado a un lado del diodo.

Para comprobar este posible cambio en la respuesta visual, determiné primero la curva de respuesta de Bruce Firtha (con el ojo izquierdo solamente) en una habitación en penumbra. Su ojo derecho estaba tapado con un trozo de cartón. Firtha miró directamente al diodo como antes. La curva resultante era casi la misma que la anterior. Después repetí el experimento con Firtha, mirando a un pequeño punto desviado tres centímetros del diodo. El diodo estaba rodeado por un cartón corriente. Este procedimiento hacía que la imagen del diodo cayera unos 4 grados fuera del centro de la fóvea y, por tanto, se colocara en una zona que tiene algunos bastones. La curva de respuesta fue la misma para los niveles de iluminación altos, pero subió mucho para los niveles más bajos.

Tal subida hace pensar que, en condiciones de poca iluminación, el retraso visual se incrementa más deprisa en las zonas de la retina exteriores a la fóvea que directamente en el centro de la misma. La visión con luz fuerte se llama fotópica, y está determinada por la respuesta de los conos. La visión con luz débil se llama escotópica y está determinada por la respuesta de los bastones.

Los resultados nos llevan a pensar que la visión escotópica fuera del centro de la fóvea produce un mayor retraso visual que en el centro de la fóvea, debido a la abundancia de bastones.

Se puede hacer otro tipo de trabajos sobre los retrasos visuales, bien con este tipo de aparato o con otro. ¿Qué trayectoria sigue la curva de respuesta en función de la distancia angular desde el centro de la fóvea? ¿Se podrían representar las curvas de respuesta de la zona que rodea a la fóvea? Si los trabajos subsiguientes indican que la zona del centro de la fóvea, que carece de bastones, tiene una curva de respuesta diferente de las curvas de otras zonas de la retina, se podría pensar que el retraso visual de los bastones es distinto del que tienen los conos.

Se pueden también poner blancos estimulantes de diferentes tamaños. Un blanco pequeño, al que se mira directa-

mente, estimula sólo el centro de la fóvea mientras que un blanco mayor incita una porción mayor de la retina y, por consiguiente, podría provocar respuesta de los bastones. La respuesta visual en este caso es una superposición sobre toda la zona de la retina que es estimulada.

No estoy muy seguro de cómo deberían ser las curvas de respuesta en función de las luces de la habitación. Aunque no he notado ningún cambio en la curva de respuesta al encender las luces (por lo menos en los datos que tomé) quizá al lector le interese repetir el experimento. Se podrían iluminar los ojos directamente con una luz pequeña en lugar de iluminar el diodo estimulador y sus entornos. A medida que los ojos se van adaptando a un nivel de iluminación superior ¿responden visiblemente más pronto? Este, al menos, fue el resultado que obtuve con el efecto del péndulo.

El lector debe recordar que en cualquier experimento se deben recoger muchos datos, promediar los resultados y comparar después éstos teniendo en cuenta las desviaciones típicas. Dos resultados medios que han sido recogidos bajo circunstancias distintas y que caen dentro de una desviación típica uno del otro, se consideran normalmente como indiferenciables.

El retraso visual puede afectar a la propia percepción del paisaje cuando se va en coche, si se mira hacia un lado de la carretera llevando puestas unas gafas de sol con cristal en un ojo solamente. Para ello, se debe fijar la vista en objetos distantes, que estén situados primero en un lado del coche y después en el otro. Al mismo tiempo se irán comparando las velocidades a que parecen pasar los objetos situados en primer plano en ambos lados. Si se tapa el ojo izquierdo con el cristal obscuro, los objetos de la izquierda parecen pasar más despacio que los de la derecha. Si se estima la velocidad del coche por el movimiento del paisaje exterior, la vista del lado izquierdo indica que el coche se está moviendo más lentamente de lo real, mientras que la del lado derecho dice lo contrario, parece que se mueve a mayor velocidad de la real.

Este procedimiento altera también los tamaños y distancias de los objetos en primer plano. Los que están del lado del coche que da sensación de menor velocidad tienden a parecer más próximos y más pequeños de lo que verdaderamente son. Los objetos del otro lado parecen estar más lejanos y ser mayores de lo que ciertamente son.

Este falseamiento de la velocidad, distancia y tamaño se produce por el retraso visual que ocasiona el filtro obscuro puesto enfrente de un ojo. Ese ojo envía una respuesta retrasada al cerebro. Como ocurría con el efecto del péndulo, la respuesta retrasada altera el ángulo entre un objeto de primer plano y el objeto de referencia situado en un plano posterior. Si es el ojo izquierdo el obstruido y se mira hacia la derecha, el ángulo entre el objeto en primer plano y el de referencia se incrementa debido al retraso. Este incremento significa que el objeto en primer plano está más lejos de lo que realmente está. Si se mira al objeto como estando más lejano aunque ocupe el mismo tamaño angular en el campo de visión, el observador se ve obligado a reconocer que es mayor de lo que realmente es. De la misma forma, si se juzga que está más distante, pero que todavía cruza el campo de visión a la misma velocidad (es decir, en una cantidad de tiempo determinada cruza un ángulo concreto del campo de visión del observador independientemente de a qué distancia crea éste que está), habrá que reconocer que se mueve a mayor velocidad de la real. El retraso visual altera la percepción de la profundidad de los objetos, a consecuencia de lo cual el observador no puede apreciar su tamaño ni la velocidad a que pasan.

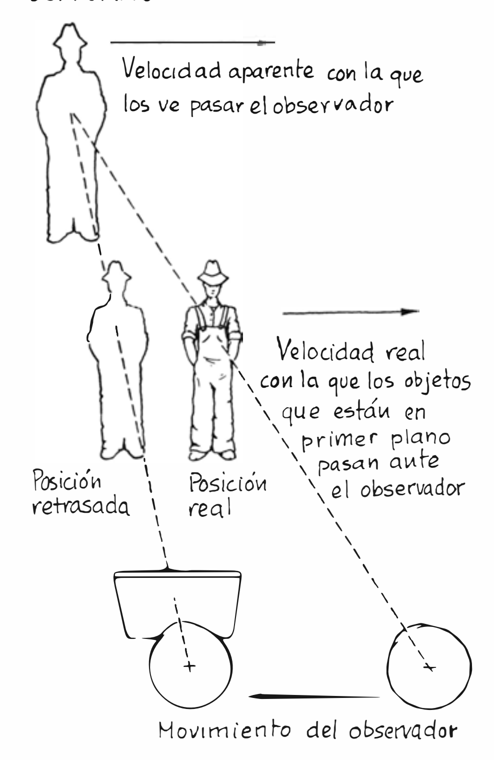
Este falseamiento de la velocidad, la distancia y el tamaño se puede demostrar ante una pequeña audiencia empleando una película hecha desde un coche en movimiento. James T. Enright, de la Universidad de California en San Diego, aconseja que la cámara enfoque algún objeto distante del lado de la carretera mientras el coche lleve una velocidad de unos 20 kilómetros por hora. El efecto quedará muy claro si existe ese objeto distante y además otros en primer plano a una distancia de unos 100 metros. Cada observador de la audiencia debe llevar gafas de sol con un solo cristal o ponerse cualquier otro filtro oscuro delante de un ojo.

Ronald Ives escribía en 1940, en el *Journal of the Franklin Institute,* que al atardecer las luces de las calles parecían encenderse ligeramente antes en los cruces que en las partes centrales. Incluso afirmaba que en el segundo grupo parecía que las luces se iban encendiendo sucesivamente a lo largo de la calle tardando más las más lejanas al observador. Sin embargo, dado que las farolas de las calles tenían un mismo circuito, el efecto tenía que ser una ilusión óptica.

Este efecto se debía al retraso visual que he estado describiendo. Había más luces en los cruces que en ningún otro punto entre éstos y, por tanto, el observador recibía más luz de los cruces. El retraso visual para un estímulo brillante es menor que para otro tenue. Aunque todas las luces se encendían simultáneamente, las de los cruces se veían antes debido a la reducción del retraso visual procedente del estímulo. Las luces a lo largo de la calle parecían irse encendiendo una por una, siendo las últimas las más lejanas. Esto era debido a que su estímulo era progresivamente más débil y ello producía un incremento en el retraso visual.

¿Puede afectar, significativamente, el retraso visual, producido por las gafas de sol, al tiempo de reacción ante un imprevisto? Es posible. Supongamos que unas gafas de sol son lo suficientemente obscuras como para añadir 100 milisegundos al tiempo de respuesta de una persona (un caso extremo). Si esa persona estuviera conduciendo a una velocidad de unos 90 kilómetros por hora (unos 25 metros por segundo), el tiempo de respuesta incrementado significaría una distancia de parada adicional de unos 2,5 metros.





Cambios registrados por un observador en movimiento

Libros

Fecundidad de la teoría general de sistemas e introducción al conductismo contemporáneo

Ramón Margalef y Ramón Bayés

El Macroscopio. Hacia una Visión Global, por Joël de Rosnay. Traducción del francés por F. Sáez Vacas. Editorial A.C., Madrid, 1977, 289 páginas. La edición original de esta obra se publicó en Francia en 1975 y su autor es Director de Aplicaciones de la Investigación, en el Instituto Pasteur, de París. Aunque el autor promete un libro de un estilo completamente nuevo, nos ofrece en realidad una serie de ensayos brillantes, siempre alrededor de un hilo conductor constituido por las relaciones entre la física y la biología, entre la ecología y la economía, con una preocupación de fondo, muy francesa, muy en la línea de Bergson y Teilhard, sobre los problemas del tiempo, de la entropía, de la emergencia del orden y de la complejidad, y de las paradojas que presenta toda evolución cuando se contempla hacia delante, o hacia atrás.

El libro se lee con gusto y debe reconocerse que es una exposición brillante e inteligente de un gran número de problemas de actualidad, que entroncan o se pueden coordinar dentro de lo que se llama teoría general de sistemas. Estos problemas se debaten en relación con la sociedad humana, lo cual da al libro actualidad en las áreas de la energía, ecología, alimentos, información, etc., sin la pesadez en que caen los ecologistas al uso, cuando se desmelenan contando miserias.

La exposición es clara, y aunque hay alguna imprecisión que comporta el estilo usado, se puede calificar de correcta. En algunas secciones es más una yuxtaposición de información que una estructura trabada, y puede aparecer algo atropellada y aun irritante para el lector que desea pensar un poco más profundamente el contenido de un libro que, en conjunto, está bien construido. Su interés por problemas políticos y sociales evita, sin embargo, tocar las desigualdades creadas y amplificadas en los intercambios de energía y ciertos problemas demográficos igualmente delicados; pero, en compensación, nos ofrece una comparación de la "actitud tradicional" y la "actitud emergente" que más parece una caracterización de buenos y malos para un western ecosociológico, que no una evaluación humanista de la encrucijada en que nuestra civilización se encuentra.

También se ocupa de la educación. Afortunadamente, el hombre es una especie capaz de soportar los más diversos y disparatados ensayos pedagógicos y seguir dando ejemplares admirables. Pero esto es una opinión del que suscribe esta reseña y su relación con lo que dice el libro es más bien tenue. En conclusión, es un libro positivo, que se lee con gusto, y que pasa revista a una serie de puntos de vista que deben tenerse en cuenta al enfrentarse con la realidad actual. (R. M.)

Introduccion al Conductismo Moderno, por H. Rachlin. Traducción de Francisco Pabón y revisión de Víctor García-Hoz Rosales. Editorial Debate, Madrid, 1977. De la misma manera que existen unas leyes naturales que regulan el proceso evolutivo de las especies y que actúan seleccionando las características de los organismos que sobrevivirán, es plausible que, a nivel ontogenético, existan otras leyes naturales que regulen las características de los comportamientos que muestran los organismos en cada momento histórico concreto. El conductismo –ciencia del comportamiento– trata de descubrir estas leyes.

Se dispone ya de alguna prueba que nos indica que tales leyes relacionan el comportamiento de los organismos con el marco ecológico –social y no social– en el que dicho comportamiento se desarrolla. De acuerdo con ellas, el medio ambiente selecciona qué comportamientos sobrevivirán –y con qué topografía, intensidad y frecuencia– y cuáles desaparecerán. Las mismas leyes nos permiten explicar tanto los comportamientos que consideramos "normales" como muchos otros que, en el seno de marcos culturales concretos, son etiquetados de "anormales" o "patológicos".

Los datos experimentales de que se dispone nos muestran que estas leyes tienen vigencia para todos y cada uno de los organismos individuales a través de una porción relativamente extensa de la escala filogenética; como mínimo, de la abeja al hombre. Este hecho justifica que, al igual que las demás ciencias biológicas, el conductismo utilice animales inferiores para la investigación de las leyes del comportamiento, aun cuando lo que en último término interese al investigador sea, precisamente, el comportamiento humano.

El término "Conductismo" suele evocar, desgraciadamente, resonancias emotivas desfavorables en algunos sectores de nuestra geografía, considerándosele sinónimo de "mecanicismo", "manipulación", "tecnología deshumanizadora", etcétera, e identificándosele, con frecuencia, con una psicología "estímulo-respuesta". Tales simplificaciones constituyen, sin duda alguna, una barrera ingenua pero poderosa que impide a muchos estudiosos adentrarse en los importantes hallazgos del conductismo moderno para poder calibrar, crítica pero desapasionadamente, su valor real y sus verdaderos límites y alcance. Por esta razón, la traducción al castellano del libro de Rachlin nos parece sumamente oportuna, ya que se trata de un texto didáctico, muy cuidado desde el punto de vista científico y que se encuentra prácticamente al día en cuanto al nivel alcanzado por la ciencia del comportamiento; la versión que comentamos corresponde a la segunda versión revisada de 1976 y en ella se mencionan ya importantes trabajos recientes como, por ejemplo, la implantación de lenguaje simbólico en chimpancés.

En *Introducción al Conductismo Moderno* se exponen, con claridad, los antecedentes, evolución, realidad y perspectivas de la línea conductista dentro de la psicología actual, siendo conveniente mencionar que, dentro de dicha línea, Rachlin, aun aludiendo a los puntos de vista de Guthrie, Hull o Tolman, da preferencia a un enfoque en buena parte skinneriano.

Aunque tal elección pueda escandalizar a algunos, en nuestra opinión ayuda a conferir a la exposición una gran agilidad y coherencia, haciendo posible que el estudiante pueda utilizar el libro como guía eficaz para adentrarse en la complicada jungla de las denominadas teorías del aprendizaje. Al contrario de los textos de Kimble (1961) y Mackintosh (1974), excelentes obras de referencia pero difícilmente digeribles por el principiante, el libro de Rachlin, alineándose junto a las excelentes obritas de Bachrach (1965) y Reynolds (1968) –cuyas ediciones castellanas son, por desgracia, difícilmente asequibles, en la actualidad, al lector español–,

constituye, sin duda, una buena plataforma para lanzarse, si se desea, a la conquista de metas más ambiciosas.

Howard Rachlin es profesor de psicología en la Universidad Estatal de Nueva York (Stony Brook), habiendo conseguido su doctorado en la Universidad de Harvard, de la que fue profesor hasta 1969. Forma parte del consejo editorial de la prestigiosa revista científica *Journal of the Experimental Analysis of Behavior* y es autor de otro libro que, en estos momentos, se está traduciendo al castellano y que se encuentra anunciado con el título de *Comportamiento y aprendizaje*.

La obra que comentamos empieza con una introducción histórica, extraordinariamente bien elaborada, que partiendo del dualismo cartesiano llega hasta la reflexología y el conductismo watsoniano. Se trata, en nuestra opinión, de una de las síntesis más breves y accesibles que conocemos sobre los antecedentes, evolución, problemas y alternativas de la psicología como ciencia.

El segundo capítulo se ocupa de las técnicas y procedimientos básicos que constituyen la característica esencial del conductismo moderno. Tras explicar con cierto detalle algunos de los experimentos de Pavlov, Rachlin pasa revista al condicionamiento instrumental y compara, luego, ambos paradigmas. Finalmente, presenta diversas técnicas que se utilizan en el laboratorio para el estudio y medición del comportamiento. En este capítulo, también se examinan algunos problemas teóricos de difícil planteamiento como, por ejemplo, el de la inhibición interna, saliendo el autor airoso de la prueba.

El tercero se inicia con la presentación de cuatro importantes teorías del reforzamiento: 1) La reducción de la necesidad; 2) La reducción de la tensión; 3) La estimulación del cerebro; 4) La respuesta como reforzamiento. Rachlin termina su revisión, breve pero substanciosa, con las siguientes palabras: "Lo importante es recordar que ninguna de ellas (de las teorías del reforzamiento) es suficiente por sí sola para explicar todos los reforzadores conocidos. Ello significa que la mejor forma de saber si un estímulo refuerza es poniéndolo a prueba en un experimento de condicionamiento instrumental" (p. 146). El capítulo prosigue con una visión panorámica de los programas de reforzamientos fundamentales, del castigo, y de los comportamientos de evitación y escape. Es especialmente clara la distinción entre reforzamiento negativo y castigo, conceptos que suelen confundirse.

El cuarto capítulo trata de las limitaciones biológicas del aprendizaje, que no aparecía en la primera edición americana de 1970. Con él, Rachlin intenta cubrir una laguna existente en su trabajo anterior, puesta de relieve por algunos de sus críticos. En efecto, incluso en las condiciones controladas del laboratorio, no es cierto que siempre que se correlacionen debidamente los acontecimientos ambientales y los conductuales se produzca condicionamiento. Existen diferencias en la receptibilidad del sujeto –ejemplificadas en el caso límite de la impronta–, diferencias en las reacciones motores y diferencias de asociabilidad, ya que las tendencias asociativas preestablecidas desempeñan su papel tanto en el condicionamiento clásico como en el instrumental.

En el siguiente capítulo se estudian las relaciones entre estímulos, respuestas y reforzamiento. Aun cuando existe dentro de la psicología experimental una importante corriente "conexionista", que busca la explicación de los aprendizajes en la comprensión de los mecanismos neurofisiológicos que tienen lugar en el organismo, existe otra corriente no menos importante, a la cual se apunta Rachlin –y a la que, personalmente, también nos adherimos–, que podríamos denominar "correlacional", la cual centra preferentemente su atención en las relaciones existentes entre el comportamiento de los organismos y su entorno. Evidentemente, ambas teorías nos colocan sobre la pista de acontecimientos inobservados; sin embargo, la diferencia fundamental entre ellas estriba en que mientras que la primera busca "conexiones" en el interior del organismo, la segunda pone su énfasis en las interacciones entre el comportamiento y su ambiente, las cuales son, en principio, asequibles al investigador y susceptibles de ser alteradas, observadas y registradas directamente: "Cuando, por ejemplo, una rata aprieta una barra para conseguir comida, centramos nuestra atención en la relación entre la comida y el apretar la palanca y no en la relación entre la comida y las esperanzas, expectativas, miedos, deseos y sueños de la rata.

Si la rata aprende una serie compleja de movimientos, centramos nuestra atención en las manipulaciones del entorno necesarias para que produzca la serie compleja de movimientos, y no en la forma en que estos movimientos son codificados en la mente de la rata o incluso en su cerebro" (p. 236).

Finalmente, en el sexto y último capítulo, Rachlin nos muestra algunas de las direcciones en las que se extiende el conductismo: las terapéuticas conductuales, el autocontrol, el comportamiento verbal, y la cuantificación de las variables de comportamiento. El libro se cierra con un glosario de términos y un índice analítico, quizá algo limitado, si bien de gran utilidad.

Para el lector español el libro presenta una dificultad especial debido a que muchas de las referencias a las que remite el libro, para ampliación complementaria, corresponden a trabajos publicados en *Journal of the Experimental Analysis of Behavior,* revista difícilmente asequible, tanto a nivel físico –a finales de 1971, de acuerdo con datos facilitados por la propia revista, el único suscriptor en la península ibérica era el autor de estas líneas–, como de comprensión. Al terminar la lectura del libro nos hemos quedado con la impresión de que Rachlin es, ante todo, un gran pedagogo. ¡Ojalá muchos de los admirables trabajos que aparecen en sus referencias tuvieran la misma claridad!

Posiblemente, habrá quien considere que el libro de Rachlin incurre en numerosas y flagrantes omisiones. Por nuestra parte, estimamos que, en el campo del aprendizaje, existen dos tipos de libro: el que podríamos denominar "libro-fotografía", el cual trata de no tomar partido y de describirnos todo lo que se conoce, desde todos los puntos de vista, sobre la región desconocida que pretendemos explorar, y el "libro-mapa", que nos señala, con claridad, unos itinerarios para que nos adentremos en ella. En el primer caso, el riesgo estriba en que ante el inmenso y complejo panorama que aparece ante nuestros ojos, se nos escape una expresión de desaliento y decidamos no emprender excursión alguna; en el segundo, siempre cabe la posibilidad de que los itinerarios que se nos ofrece no sean los mejores y que nos perdamos la estupenda vista de las cataratas. Entre ambas opciones, aun a riesgo de equivocarnos, siempre hemos preferido la segunda. Aparte del confusionismo que suelen crear muchos de los libros del primer tipo, como dice Bachrach, "ser ecléctico puede significar que se tienen los pies firmemente asentados en el aire".

Para concluir, resumiremos diciendo que, en nuestra opinión, *Introducción al Conductismo Moderno* es un libro excelente, sumamente recomendable para cursos de introducción a la psicología, o, simplemente, para aquellas personas que deseen conocer algunas de las características distintivas del conductismo de los años setenta y no solo del de los años veinte, como suele ser la tónica en muchos manuales que gozan de amplia difusión entre nosotros. (R. B.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes;

LA PRODUCCION MUNDIAL DE PETROLEO

PETROLEUM RESOURCES: HOW MUCH OIL AND WHERE. John D. Moody y Robert E. Geiger en *Technology Review,* págs. 38-45; marzo/abril, 1975.

ENERGY: GLOBAL PROSPECTS 1985-2000: Workshop on Alternative Energy Strategies. McGraw-Hill Book Company, 1977.

ENERGY SUPPLY TO THE YEAR 2000: GLOBAL AND NATIONAL STUDIES. Workshop on Alternative Energy Strategies. The MIT Press, junio, 1977.

LEPTONES PESADOS

ELECTRON-POSITRON ANNIHILATION AND THE NEW PARTICLES. Sidney D. Drell en *Scientific American,* vol. 232, n.º 6, págs. 50-62; junio, 1975.

WHAT IS THE WORLD MADE OF? Gerald Feinberg. Anchor Press, 1977.

PARTICULAS FUNDAMENTALES CON ENCANTO. Roy F. Schwitters en *Investigación y Ciencia,* n.º 15, págs. 32-48; diciembre, 1977.

LA SUPERFICIE DE MARTE

MARS AS VIEWED BY MARINER 9: NASA SPECIAL PUBLICATION 329. National Aeronautics and Space Administration. Government Printing Office, 1974.

THE NEW MARS: THE DISCOVERIES OF MARINER 9. William K. Hartmann y Odell Raper. NASA Special Publication 337. Government Printing Office, 1974.

MARS. James B. Pollack en *Scientific American,* vol. 233, n.º 3, págs 107-117; septiembre, 1975.

THE GEOLOGY OF MARS. Thomas A. Mutch, Raymond E. Arvidson, James W. Head, Kenneth L. Jones y R. Stephan Saunders. Princeton University Press, 1977.

SCIENTIFIC RESULTS OF THE VIKING PROJECT. *Journal of Geophysical Research,* vol. 82, n.º 28; 30 de septiembre de 1977.

EL FLUJO DE ENERGIA EN UN ECOSISTEMA DE BOSQUE

ENERGY EXCHANGE IN THE BIOSPHERE. D. M. Gates. Harper & Row, 1962.

ECOLOGY. Robert E. Ricklefs. Chiron Press, 1973.

BIOCHEMISTRY OF A FORESTED ECOSYSTEM. Gene E. Likens, F. Herbert Bormann, R. S. Pierce, J. S. Eaton y N. M. Johnson. Springer-Verlag, 1977.

COMO FABRICAN ATP LAS CELULAS

PROTON-TRANSLOCATION PHOSPHORYLATION IN MITOCHONDRIA, CHLOROPLASTS AND BACTERIA: NATURAL FUEL CELLS AND SOLAR CELLS. Peter Mitchell en *Federation Proceedings,* vol. 26, n.º 5, págs. 1370-1379; septiembre-octubre, 1967.

KINETICS OF ADENOSINE TRIPHOSPHATE SYNTHESIS IN BOVINE HEART SUBMITOCHONDRIAL PARTICLES. William S. Thayer y Peter C. Hinkle en *The Journal of Biological Chemistry,* vol. 250, n.º 14, págs. 5336-5342; 25 de julio de 1975.

A NEW LOOK AT MECHANISMS IN BIOENERGETICS. Efraim Racker. Academic Press, Inc., 1976.

OXIDATIVE PHOSPHORYLATION AND PHOTOPHOSPHORYLATION. Paul D. Boyer, Britton Chance, Lars Ernster, Peter Mitchell, Efraim Racker y E. C. Slater en *Annual Review of Biochemistry,* vol. 46, págs. 955-1026; 1977.

MATEMATICA COMBINATORIA DE LA PLANIFICACION

ON PACKING SQUARES WITH EQUAL SQUARES. Paul Erdös y Ronald L. Graham. *Journal of Combinatorial Theory (Series A),* vol. 19, n.º 1, págs. 119-123; julio, 1975.

COMPUTER AND JOB-SHOP SCHEDULING THEORY. Dirigido por Edward G. Coffman. John Wiley & Sons, Inc., 1976.

NEW DIRECTIONS IN CRYPTOGRAPHY. Whitfield Diffie y Martin E. Hellman en *IEEE Transactions on Information Theory,* vol. IT-22, n.º 6, págs. 644-654; noviembre, 1976.

BRUEGEL EL VIEJO, GUIA PARA EL ESTUDIO DE LA CIENCIA DEL s. XVI

GRAPHIC WORLDS OF PETER BRUEGEL THE ELDER. H. Arthur Klein. Dover Publications, Inc., 1963.

GREAT STRUCTURES OF THE WORLD. H. Arthur Klein y Mina C. Klein. The World Publishing Company, 1968.

PETER BRUEGEL THE ELDER, ARTIST OF ABUNDANCE: AN ILLUSTRATED PORTRAIT OF HIS LIFE, ERA, AND ART. H. Arthur Klein y Mina C. Klein. The Macmillan Company, 1968.

SCIENCE AND SOCIETY IN THE SIXTEENTH AND SEVENTEENTH CENTURIES. Alan G. R. Smith. Thames and Hudson, Londres, 1972.

THE WORLD OF MEASUREMENTS. H. Arthur Klein. Simon and Schuster, 1974.

EL MANTENIMIENTO DE LA VARIABILIDAD GENETICA DE LAS POBLACIONES

GENETICS OF THE EVOLUTIONARY PROCESS. Theodosius Dobzhansky. Columbia University Press. 1971.

THE GENETIC BASIS OF EVOLUTIONARY CHANGE. Richard C. Lewontin. Columbia University Press. 1974.

MOLECULAR EVOLUTION. Dirigido por Francisco J. Ayala. Sinauer Associates, Inc. 1976.

JUEGOS MATEMATICOS

COMPUTABILITY AND UNSOLVABILITY. Martin Davis. McGraw-Hill Book Company, 1958.

GÖDEL'S PROOF. Ernest Nagel y James R. Newman. New York University Press, 1958.

THE UNDECIDABLE. Dirigido por Martin Davis. Raven Press. 1965.

TALLER Y LABORATORIO

STEREOPSIS, VISUAL LATENCY, AND THREE-DIMENSIONAL MOVING PICTURES. James T. Enright en *American Scientist,* vol. 58, n.º 5, págs. 536-545: septiembre-octubre, 1970.

LATENCY FUNCTIONS IN HUMAN VISION. R. J. W. Mansfield en *Vision Research,* vol. 13, págs. 2219-2234; 1973.

EYE MOVEMENTS AND THE PULFRICH PHENOMENON, B. J. Rogers. J. J. Steinbach y H. Ono en *Vision Research,* vol. 14, págs. 181-185, 1974.